

**STIC-Biotech/ChemLib**

---

**From:** Seharaseyon, Jegatheesan  
**Sent:** Thursday, July 12, 2001 10:37 AM  
**To:** STIC-Biotech/ChemLib  
**Subject:** 09/554,933

Please search SEQ ID No: 3 of 09/554,933. Thanks.

J.Seharaseyon  
Art Unit 1647  
CM1 10D16  
10CO1 MB  
(703)-305-1112

**THIS PAGE BLANK (USPTO)**

## Scientific and Technical Information Center

**If more than one search is submitted, please prioritize searches in order of need.**

\*\*\*\*\*

Title of Invention: \_\_\_\_\_

Inventors (please provide full names): \_\_\_\_\_

Earliest Priority Filing Date: \_\_\_\_\_

*\*For Sequence Searches Only\* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

~~STAFF USE ONLY~~

Searcher: W. H. H. - 104

Searcher Phone #: 308-45010

Searcher Location: Grated lot.

Date Searcher Picked Up: 7/12/01

Date Completed: 7/16/01

Searcher Prep &amp; Review Time:

Clerical Prep Time: 3 min.

Online Time: 2m

### Type of Search

NA Sequence (#)\_\_\_\_\_

AA Sequence (#) 1

Structure (#) \_\_\_\_\_

Bibliographic \_\_\_\_\_

**Litigation** .....

Fulltext

Patent Family

Other \_\_\_\_\_

**Vendors and cost where applicable**

STN \_\_\_\_\_

Dialog \_\_\_\_\_

Questel:Orbit \_\_\_\_\_

Dr.Link \_\_\_\_\_

Lexis/Nexis \_\_\_\_\_

Sequence Systems AB5502

WWW/Internet \_\_\_\_\_

Other (specify) \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 11:38:20 ; Search time 20.2 Seconds  
(without alignments)  
1149.455 Million cell updates/sec

Title: US-09-554-933-3

Perfect score: 2080  
Sequence: 1 MAGIPGLFLFLFLCAVQ.....LKAYOICWIKGNLYDCREG 383

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Archived: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: /SIDSR/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDSR/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDSR/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDSR/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDSR/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDSR/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDSR/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDSR/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDSR/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDSR/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDSR/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDSR/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDSR/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDSR/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDSR/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDSR/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDSR/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDSR/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDSR/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDSR/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDSR/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDSR/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	20	AAV08657 Human transmembran
2	2080	100.0	383	20	AAV08650 MO9927094 Seq ID 1
3	2080	100.0	383	20	AAV13390 Amino acid sequenc
4	2080	100.0	383	21	AAZ25592 Protein encoded by
5	2080	100.0	383	21	AAZ25618 Protein encoded by
6	2080	100.0	383	21	AAV88277 Human TANGO 186 pr
7	2080	100.0	383	21	AAV87270 Human signal pepti
8	2080	100.0	383	21	AAV53627 A bone marrow secr
9	2080	100.0	383	22	AAAB80258 Human PRO307 prote
10	2080	100.0	383	22	AAAB48974 Human Zsig13 varia
11	2080	100.0	392	22	AAAB48973 Human Zsig13 varia

12	2073	99.7	383	22	AAAB8336 Human membrane or
13	2044	98.3	392	22	AAAB48972 Human Zsig13 varia
14	1909.5	91.8	375	21	AAAB25585 Protein encoded by
15	1909.5	91.8	413	21	AAAB25617 Protein encoded by
16	1890.5	90.9	382	21	AAV88288 Murine TANGO 186 p
17	1041	50.0	413	20	AAV41728 Human PRO1057 prot
18	1041	50.0	413	20	AAV28681 Human PRO1057 secr
19	1041	50.0	413	21	AAAB4284 Human PRO1057 (UNQ
20	202.5	9.7	316	13	AAAE00013 Bacillus lichenifo
21	194	9.3	316	13	AAAE00011 Protease Blase. S
22	193	9.3	316	22	AAAE00011 Bacillus lichenifo
23	184	8.8	222	22	AAAE00114 Bacillus lichenifo
24	183	8.8	222	22	AAAE00040 Bacillus lichenifo
25	183	8.8	222	22	AAAE00041 Bacillus lichenifo
26	183	8.8	222	22	AAAE00089 Bacillus lichenifo
27	183	8.8	222	22	AAAE00090 Bacillus lichenifo
28	183	8.8	222	22	AAAE00193 Bacillus lichenifo
29	181	8.7	222	22	AAAE00080 Bacillus lichenifo
30	181	8.7	222	22	AAAE00108 Bacillus lichenifo
31	181	8.7	222	22	AAAE00163 Bacillus lichenifo
32	181	8.7	222	22	AAAE00169 Bacillus lichenifo
33	180	8.7	222	22	AAAE00037 Bacillus lichenifo
34	180	8.7	222	22	AAAE00073 Bacillus lichenifo
35	180	8.7	222	22	AAAE00115 Bacillus lichenifo
36	180	8.7	222	22	AAAE00188 Bacillus lichenifo
37	179	8.6	222	22	AAAE00059 Bacillus lichenifo
38	179	8.6	222	22	AAAE00076 Bacillus lichenifo
39	179	8.6	222	22	AAAE00082 Bacillus lichenifo
40	179	8.6	222	22	AAAE00091 Bacillus lichenifo
41	179	8.6	222	22	AAAE00166 Bacillus lichenifo
42	179	8.6	222	22	AAAE00170 Bacillus lichenifo
43	179	8.6	222	22	AAAE00172 Bacillus lichenifo
44	179	8.6	222	22	AAAE00180 Bacillus lichenifo
45	179	8.6	222	22	AAAE00190 Bacillus lichenifo

## ALIGNMENTS

RESULT 1	
ID	AAV08657 standard; Protein: 383 AA.
XX	AAV08657:
XX	09-AUG-1999 (first entry)
DE	Human transmembrane domain containing protein from clone HP10493.
XX	
KW	Transmembrane domain; human; nutrition; cytokine; cell differentiation;
KW	immune stimulation; immune suppression; haematopoiesis; activin;
KW	regulatory tissue growth; inhibin; chemostatic; chemokinetic;
KW	haemostatic; thrombolytic; tumour inhibitor; anti-inflammatory;
KW	gene therapy; screening.
XX	
OS	Homo sapiens.
XX	
PN	MO9927094-A2.
PD	03-JUN-1999.
XX	
PF	20-NOV-1998; 98MO-JF05238.
XX	
PR	25-NOV-1997; 97JP-0323129.
XX	
PA	(PROT-) PROTEGENE INC.
PA	(SAGA) SAGAMI CHEM RES. CENT.
XX	
PI	Kato S, Kimura T, Sekine S;
XX	
DR	WPI, 1999-357835/30.
DR	N-PSDB; AAX77690, AAX77691.
XX	

PT Novel proteins containing transmembrane domains, useful as  
 PT anti-inflammatories, immune stimulators/suppressors and tissue  
 growth compounds

XX Claim 1: Page 68-69; 89pp; English.

CC This invention describes novel human transmembrane containing proteins  
 CC and their encoding nucleic acids. Although no specific use is given for  
 CC the proteins, they may have a range of activities selected from  
 CC nutritional uses, cytokine and cell differentiation, immune  
 CC stimulation/suppression, haematopoiesis regulation, tissue growth,  
 CC activin/inhibin, chemostatic/chemokinetic, haemostatic/thrombolytic,  
 CC receptor/ligand, tumour inhibitor, anti-inflammatory and other undefined  
 CC as gene sources for gene therapy. The cDNAs can also be used for large  
 CC scale expression of proteins. The cDNAs can also be used for large  
 CC detection of the corresponding ligands and for screening of novel  
 CC low-molecular pharmaceuticals.

XX Sequence 383 AA;

Query Match

Best Local Similarity 100.0%; Score 2080; DB 20; Length 383;  
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLFLFLFLLCAVGOVSPYSAPWKPMPAYRLPVYLPOSTINIAKPDGAEAKLE 60  
 DB 1 magipglflflfllcavgvgsypsapwkpmpayrlpvlpstlniakpdtgaekle 60  
 QY VSSSCGPOCHGKPTPLPYEAKOYLSEYELIYANGSRTETOVGIYILSSSGDAQHSDGS 120  
 DB 61 vssscgpcgchktplpyeakqylsyelIyangsrtetovgiylsssgdaqhrdsqs 120  
 QY 121 SGSRKRRKQIYGYDSRFSIFGKDFLNTPFSTSVKLSCTGCTVAEKHVLTAAHCIDHG 180  
 DB 121 sgksrrkrqiygydsrfsifgkdfllnypfstsvklsctgctvlaekhvltaaahcldhg 180  
 QY 181 KTYVKGTOKLRVGFELKPKRKGANDSTAMPQMKFOWIRVKTVPKGMKGNAND 240  
 DB 181 ktyvkgtdglrvgflkpkrkdgandstampsqmkfgyirvkrthvpgkgnand 240  
 QY 241 TGMIDYALIELLKKPRKFKMIGVSPPAKOLPGGRHFSGYDNDRCNLYRRCVDKDE 300  
 DB 241 tgmidyalllelkpkrkfkmgivspakolpggrhfsydnndrcnlyrrcfdvkd 300  
 QY 301 TYDLTYOCCDAOPGASGSGVYVYRMKROQKWERKIIIGIFSGHQWVDMNGSPDDFNVAVR 360  
 DB 301 tydllyqcdadpagsgsgvyvymkrqgkwerkiiigifsgbqwdmngspddfnvavr 360  
 QY 361 TTPLYAOTICWIKGNVLDCCREG 383  
 DB 361 ttplyaqicwyikgnvldcreg 383

RESULT 2

ID AAY08660 standard; Protein; 383 AA.

AC AAY08660;

DT 09-AUG-1999 (first entry)

DE WO9927094 Seq ID 12.

XX Transmembrane domain; human; nutrition; cytokine; cell differentiation;  
 KW Immune stimulation; immune suppression; haematopoiesis; activin;  
 KW regulatory tissue growth; inhibin; chemostatic; chemokinetic;  
 KW haemostatic; thrombolytic; tumour inhibitor; anti-inflammatory;  
 KW gene therapy; screening.

OS Homo sapiens.

PN WO9927094-A2.

XX 03-JUN-1999.

PF 20-NOV-1998; 98MO-JP05238.

XX 25-NOV-1997; 97JP-0323129.

PA (PROT-) PROTEGENE INC.

XX (SAGA) SAGAMI CHEM RES CENT.

PI Kato S, Kimura T, Sekine S;

XX WPI; 1999-357835/30.

PT Novel proteins containing transmembrane domains, useful as  
 PT anti-inflammatories, immune stimulators/suppressors and tissue  
 growth compounds

PS Disclosure; Page 87-89; 89pp; English.

CC This invention describes novel human transmembrane containing proteins  
 CC and their encoding nucleic acids. Although no specific use is given for  
 CC the proteins, they may have a range of activities selected from  
 CC nutritional uses, cytokine and cell differentiation, immune  
 CC stimulation/suppression, haematopoiesis regulation, tissue growth,  
 CC activin/inhibin, chemostatic/chemokinetic, haemostatic/thrombolytic,  
 CC receptor/ligand, tumour inhibitor, anti-inflammatory and other undefined  
 CC as gene sources for gene therapy. The cDNAs can also be used for large  
 CC scale expression of proteins. The cDNAs can also be used for large  
 CC detection of the corresponding ligands and for screening of novel  
 CC low-molecular pharmaceuticals.

SO Sequence 383 AA;

Query Match

Best Local Similarity 100.0%; Score 2080; DB 20; Length 383;  
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLFLFLFLLCAVGOVSPYSAPWKPMPAYRLPVYLPOSTINIAKPDGAEAKLE 60  
 DB 1 magipglflflfllcavgvgsypsapwkpmpayrlpvlpstlniakpdtgaekle 60  
 QY VSSSCGPOCHGKPTPLPYEAKOYLSEYELIYANGSRTETOVGIYILSSSGDAQHSDGS 120  
 DB 61 vssscgpcgchktplpyeakqylsyelIyangsrtetovgiylsssgdaqhrdsqs 120  
 QY 121 SGSRKRRKQIYGYDSRFSIFGKDFLNTPFSTSVKLSCTGCTVAEKHVLTAAHCIDHG 180  
 DB 121 sgksrrkrqiygydsrfsifgkdfllnypfstsvklsctgctvlaekhvltaaahcldhg 180  
 QY 181 KTYVKGTOKLRVGFELKPKRKGANDSTAMPQMKFOWIRVKTVPKGMKGNAND 240  
 DB 181 ktyvkgtdglrvgflkpkrkdgandstampsqmkfgyirvkrthvpgkgnand 240  
 QY 241 TGMIDYALIELLKKPRKFKMIGVSPPAKOLPGGRHFSGYDNDRCNLYRRCVDKDE 300  
 DB 241 tgmidyalllelkpkrkfkmgivspakolpggrhfsydnndrcnlyrrcfdvkd 300  
 QY 301 TYDLTYOCCDAOPGASGSGVYVYRMKROQKWERKIIIGIFSGHQWVDMNGSPDDFNVAVR 360  
 DB 301 tydllyqcdadpagsgsgvyvymkrqgkwerkiiigifsgbqwdmngspddfnvavr 360  
 QY 361 TTPLYAOTICWIKGNVLDCCREG 383  
 DB 361 ttplyaqicwyikgnvldcreg 383

RESULT 3

ID AAY13390 standard; Protein; 383 AA.

XX AAV13390;  
AC 25-JUN-1999 (first entry)  
DT  
XX Amino acid sequence of protein PRO307.  
DE  
XX Secreted protein; transmembrane protein; human; enterocolitis;  
KW Zollinger-Ellison syndrome; gastrointestinal ulceration;  
KW congenital microvillus atrophy; skin disease; cell growth;  
KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
KW fibromodulin; dermal scarring; Usher Syndrome; Atrophla areata;  
XX anti-thrombotic; wound healing; tissue repair.  
OS Homo sapiens.  
XX MO9914328-A2.  
ND 25-MAR-1999.  
XX 16-SEP-1998; 98WO-US19330.  
PE 25-NOV-1997; 97US-0066840.  
XX 17-SEP-1997; 97US-0059113.  
PR 17-SEP-1997; 97US-0059115.  
PR 17-SEP-1997; 97US-0059117.  
PR 17-SEP-1997; 97US-0059119.  
PR 17-SEP-1997; 97US-0059121.  
PR 17-SEP-1997; 97US-0059122.  
PR 17-SEP-1997; 97US-0059184.  
PR 18-SEP-1997; 97US-0059266.  
PR 18-SEP-1997; 97US-0059266.  
PR 15-OCT-1997; 97US-0062125.  
PR 17-OCT-1997; 97US-0062285.  
PR 17-OCT-1997; 97US-0062287.  
PR 21-OCT-1997; 97US-0063486.  
PR 24-OCT-1997; 97US-0062814.  
PR 24-OCT-1997; 97US-0062816.  
PR 24-OCT-1997; 97US-0063045.  
PR 24-OCT-1997; 97US-0063120.  
PR 24-OCT-1997; 97US-0063121.  
PR 24-OCT-1997; 97US-0063127.  
PR 24-OCT-1997; 97US-0063128.  
PR 27-OCT-1997; 97US-0063329.  
PR 27-OCT-1997; 97US-0063327.  
PR 28-OCT-1997; 97US-0063541.  
PR 28-OCT-1997; 97US-0063542.  
PR 28-OCT-1997; 97US-0063544.  
PR 28-OCT-1997; 97US-0063549.  
PR 28-OCT-1997; 97US-0063550.  
PR 28-OCT-1997; 97US-0063564.  
PR 29-OCT-1997; 97US-0063435.  
PR 29-OCT-1997; 97US-0063704.  
PR 29-OCT-1997; 97US-0063732.  
PR 29-OCT-1997; 97US-0063738.  
PR 29-OCT-1997; 97US-0063734.  
PR 29-OCT-1997; 97US-0064215.  
PR 29-OCT-1997; 97US-0063735.  
PR 31-OCT-1997; 97US-0063870.  
PR 31-OCT-1997; 97US-0064103.  
PR 03-NOV-1997; 97US-0064248.  
PR 07-NOV-1997; 97US-0064809.  
PR 12-NOV-1997; 97US-0065186.  
PR 17-NOV-1997; 97US-0065846.  
PR 18-NOV-1997; 97US-0065693.  
PR 21-NOV-1997; 97US-0066120.  
PR 21-NOV-1997; 97US-0066364.  
PR 24-NOV-1997; 97US-0066772.  
PR 24-NOV-1997; 97US-0066466.  
PR 24-NOV-1997; 97US-0066770.  
PR 24-NOV-1997; 97US-0066511.  
PR 24-NOV-1997; 97US-0066453.

XX (GETH ) GENTECH INC.  
PA  
XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;  
PI WPI: 1999-229533/19.  
XX DR N-PSDB; AAX52261.  
XX New isolated human genes and polypeptides used in, e.g. treatment of  
PT gastrointestinal ulceration  
XX Claim 12; Fig 96; 320pp; English.  
XX AAV13344-403 represent secreted and transmembrane human proteins.  
CC The cDNA sequences are obtained from cDNA libraries, prepared from  
CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.  
CC The encoded polypeptides have specific uses based on their homology to  
CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders  
CC associated with the preservation and maintenance of gastrointestinal  
CC mucosa and the repair of acute and chronic mucosal lesions  
CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal  
CC ulceration and congenital microvillus atrophy), skin diseases associated  
CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial  
CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),  
CC potent effects on cell growth and development, diseases related to  
CC growth or survival of nerve cells including Parkinson's disease,  
CC Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as  
CC as a target for anti-tumor drugs. PRO533 may be used in the treatment  
CC of Usher Syndrome or Atrophla areata; PRO269 can be used as an  
CC anti-thrombotic agent; PRO287 polypeptides and portions may have  
CC therapeutic applications in wound healing and tissue repair; PRO317 can  
CC be used for treating problems of the kidney, uterus, endometrium, blood  
CC vessels, or related tissue, e.g. in the heart of genital tract.  
XX  
SQ Sequence 383 AA;

Query Match 100.0%; Score 2080; DB 20; Length 383;  
Best Local Similarity 100.0%; Pred. No. 8e-147;  
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIGLLFLLFELLCAGVGVSPYSAPMKPTWPAVRLPVLPQSTLNIAKDPFGAEAKLE 60  
DB 1 magigllfllfllcavgvspysapkwlpwvayrlpvlpqstlniakdpfgaeakle 60  
QY 61 VSSSCGPOCHGTPRLPTDEAKQVLSYETLVANGSRITFOVCVITLSSGCGAQRDSCS 120  
DB 61 vssscgpgchkgtpripyleeakqylsyetlvangsrirtetvqyilsssgdaqqrds 120  
QY 61 VSSSCGPOCHGTPRLPTDEAKQVLSYETLVANGSRITFOVCVITLSSGCGAQRDSCS 120  
DB 61 vssscgpgchkgtpripyleeakqylsyetlvangsrirtetvqyilsssgdaqqrds 120  
QY 121 SGKSRKROIGYDPSRFSIFGKDFLLNYPFSTSVLSTGCTGTVAEKHYVLAACIHG 180  
DB 121 sgkstrkrqiygdsrfsifgkdfllnypfstsvlsgctgtvaeahvllaachidg 180  
QY 181 KTYVGTGKLRVGLTKPRFKDGRGANDSTAMPQMKFQWIRVKNRTHVPGWIKGNAND 240  
DB 181 ktyvgtklrvgltkprfkdggrgandstampeqmkfqrwvrtlvprgwlkgnand 240  
QY 241 IGMDYVALLELKKRHKRRKPMKIGVSPAKOLPGRIHPSGDNDRPNLTVRFDVDE 300  
DB 241 igmdyvallelkkprkrrkfmklyvspakqlpggrlhpsgdndrpnltvrrfdvde 300  
QY 301 TYDLLYQOCDAOPGASGSGVYVRMWRKROOQKWERKIIIGFSGHGVNDNGSPDFNVAVR 360  
DB 301 tydllygqcdapgsagsgvyvrwmkrrooqkwerkllifsgghvndngspdfnva 360  
QY 361 ITPLKYAOTCIWIKGNVLDCEG 383  
DB 361 itplkyaqicwyikgnylidcreg 383

RESULT 4  
AAB25592

ID AAB25592 standard; Protein; 383 AA.  
 AC AAB25592;  
 DT 21-NOV-2000 (first entry)  
 DE Protein encoded by human secreted protein gene #10 clone HUS005.  
 XX  
 XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;  
 XX antineumatic; dermatological; antiproliferative; antiarteriosclerotic;  
 XX anticancer; vulnery; antiviral; antibacterial; antifungal;  
 XX immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;  
 XX multiple sclerosis; inflammatory disorder; inflammatory bowel disease;  
 XX Crohn's disease; nephritis; hyperproliferative disorder;  
 XX cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;  
 XX melanoma; lymphoma; wound healing; human; chromosome 12.  
 OS Homo sapiens.  
 XX  
 XX MO200029435-A1.  
 XX  
 XX 25-MAY-2000.  
 XX  
 XX 27-OCT-1999; 99WO-US25031.  
 XX  
 XX 28-OCT-1998; 98US-0105971.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;  
 XX Greene JM;  
 XX WPI: 2000-387742/33.  
 XX  
 XX Isolated nucleic acid molecules encoding human secreted proteins are  
 XX used for the prevention, amelioration and treatment of autoimmune,  
 XX inflammatory, hyperproliferative and cardiovascular disorders, cancer,  
 XX wounds, and infectious diseases -  
 XX  
 XX Claim 1; Page 684-685; 803pp; English.  
 XX  
 XX The present invention relates to 12 secreted human proteins and the  
 XX nucleotide sequences encoding them. The polynucleotide sequences given  
 XX in AA80606-80623 encode the 12 secreted protein sequences given in  
 XX AA825576-B25593. The human secreted proteins have various activities  
 XX dependent on the tissues in which they are expressed. Examples of the  
 XX activities of the proteins include: immunosuppressant;  
 XX anti-inflammatory; antiarthritic; antineumatic; dermatological;  
 XX antiproliferative; antiarteriosclerotic; anticancer; vulnery;  
 XX antiviral; antibacterial; and antifungal activity. The proteins,  
 XX polypeptides, agonists and antagonists may be used to treat prevent  
 XX and/or diagnose various disease, disorders and conditions examples of  
 XX which include: immune disorders e.g. Addison's disease, rheumatoid  
 XX arthritis, dermatitis, and multiple sclerosis; inflammatory disorders  
 XX e.g. inflammatory bowel disease, Crohn's disease and nephritis;  
 XX hyperproliferative disorders such as paraproteinemias and purpura;  
 XX cardiovascular disorders e.g. coronary arteriosclerosis and myocarditis;  
 XX cancer e.g. melanoma and lymphoma. The proteins and polynucleotide  
 XX sequences may also be used in wound healing and the treatment of  
 XX infectious diseases. The human secreted protein gene #10 and protein  
 XX sequences are represented in sequences AAA0615 and AAB25585. Secreted  
 XX protein gene #10 is located on chromosome 12. Sequences AAB25616-B25618  
 XX represent alternative secreted protein #10 sequences and AA80669-80676  
 XX represent genes which are related to the secreted protein gene#10.  
 XX  
 XX Sequence 383 AA;  
 XX  
 XX Query Match 100.0%; Score 2080; DB 21; Length 383;  
 XX Best Local Similarity 100.0%; Pred. No. 8e-147;  
 XX Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX 1 MAGIPGLLEFLICAVGVSPYAPWKPTWPAVRLPVVLPGSTLNLAKPDFGAELKLE 60

Db 1 magipgllefllicavgvspysapwkwtpavryrlpvvlpgstlnlakpdfgaeakle 60  
 QY 61 VSSSCGPGCHKGTPLPTYEAKQYLSYETLLYANGSTETQVGIYILSSGCDGAQHRDGS 120  
 Db 61 vssscgpgchkgtplptyeakqylsyetllyangstetqvgiylissgddgaqhrdgs 120  
 QY 121 SGKSRKROITGYDSRFSIFGKDFLLNPFSTSVKLTGCTGLVAEKHLVAACIHG 180  
 Db 121 sgksrkroitgydsrfsifgkdfllnpyfstsrvkstgctglvaekhlvaacihdg 180  
 QY 181 KTYVKGSTQKLKRGELPKRKEDGSGANDSTSAPEQMKQWIRVKTTHPKWIGINAD 240  
 Db 181 ktyvkgstqlrygflkpkfkddgrgandstsapeqmkqwlrvkthpkwlgynand 240  
 QY 241 TGDYVYALLLELKKPKRREMKIGVSPPAKOLPGGRTHFSGVDNDRPGLVYRFGVKDE 300  
 Db 241 tgdvyalllelkkpkrrmkigvspakqlpggrthfshsgvndnrdpnlvyrfgdvkde 300  
 QY 301 TYDLTYQOCDAQPGASGSGVYRMRKROOKMERKIIIGFSGHWYDMNGSPQDFNVAVR 360  
 Db 301 tydllyqocdaqpgasgsgvyvrmrkrgqkwerkligfshgwdmngspqdfnvaivr 360  
 QY 361 ITPLKTAQICWIKNYLDGREG 383  
 Db 361 itplkyaqicwyiknyldcreg 383

RESULT 5  
 AAB25618  
 ID AAB25618 standard; Protein; 383 AA.  
 XX  
 XX AAB25618;  
 XX  
 XX 21-NOV-2000 (first entry)  
 XX  
 XX Protein encoded by human secreted protein gene #10.  
 XX  
 XX Secreted protein; immunosuppressant; anti-inflammatory; antiarthritic;  
 XX antineumatic; dermatological; antiproliferative; antiarteriosclerotic;  
 XX anticancer; vulnery; antiviral; antibacterial; antifungal;  
 XX immune disorder; Addison's disease; rheumatoid arthritis; dermatitis;  
 XX multiple sclerosis; inflammatory disorder; inflammatory bowel disease;  
 XX Crohn's disease; nephritis; hyperproliferative disorder;  
 XX cardiovascular disorder; coronary arteriosclerosis; myocarditis; cancer;  
 XX melanoma; lymphoma; wound healing; human; chromosome 12.  
 XX  
 XX Homo sapiens.  
 XX  
 XX MO200029435-A1.  
 XX  
 XX 25-MAY-2000.  
 XX  
 XX 27-OCT-1999; 99WO-US25031.  
 XX  
 XX 28-OCT-1998; 98US-0105971.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Ni J, Ruben SM, Olsen HS, Young PE, Kenny JJ, Moore PA, Wei Y;  
 XX Greene JM;  
 XX WPI: 2000-387742/33.  
 XX  
 XX Isolated nucleic acid molecules encoding human secreted proteins are  
 XX used for the prevention, amelioration and treatment of autoimmune,  
 XX inflammatory, hyperproliferative and cardiovascular disorders, cancer,  
 XX wounds, and infectious diseases -  
 XX  
 XX Disclosure; Page 169; 803pp; English.  
 XX  
 XX The present invention relates to 12 secreted human proteins and the  
 XX nucleotide sequences encoding them. The polynucleotide sequences given





Db 241 Igm dyd yall e l k k p h k r k m k i g v s p p a k q l p g r i h s y g n d n p g n l y v r f c v k d e 300  
 QY 301 T Y D L L Y O C C D A O P G A S G S G Y Y V M M K R Q O O K W E R K I I G I F S G H W D M N G S P D E F N V A V R 360  
 Db 301 t y d l l y g q c d a g p a s g s g y y v m m k r g q k w e r k i i g i f s g h w d m n g s p d e f n v a v r 360  
 QY 361 I T P L K Y A Q I C Y W I K N G N L D C R E G 383  
 Db 361 i t p l k y a q i c y w i k n g n l d c r e g 383  
 RESULT 7  
 ID AAY87270 standard; Protein; 383 AA.  
 AC AAY87270;  
 DT 11-MAY-2000 (first entry)  
 DE Human signal peptide containing protein HSP-47 SEQ ID NO:47.  
 Human signal peptide containing protein HSP-47 SEQ ID NO:47.  
 Human: signal peptide-containing protein; HSP; diagnosis; cancer;  
 inflammation; cardiovascular disease; anticancer; anti-inflammatory;  
 antimicrobial; neurotrophic; neuroprotective; cardiovascular; hepatotropic;  
 antiasthmatic; gene therapy; cell proliferation; neurological disorder;  
 reproductive disorder; developmental disorder; arteriosclerosis;  
 cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;  
 asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;  
 Parkinson's disease; Huntington's diseases; ovulatory defect;  
 muscular dystrophy.  
 KW Homo sapiens.  
 OS  
 XX  
 PN WO200000610-A2.  
 PD 06-JAN-2000.  
 XX  
 PF 25-JUN-1999; 99WO-US14484.  
 XX  
 PR 26-JUN-1998; 98US-0090762.  
 PR 31-JUL-1998; 98US-0094983.  
 PR 01-OCT-1998; 98US-0102686.  
 PR 11-DEC-1998; 98US-0112129.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 XX  
 PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;  
 PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;  
 PI Bandman O;  
 DR WPI: 2000-160673/14.  
 N-PSDB; AAY98155.  
 XX  
 PT New human signal peptide-containing protein useful in treatment,  
 PT prevention and diagnosis of e.g. cancer, inflammation and  
 PT cardiovascular disease  
 XX  
 PS Claim 1: Page 192-193; 327pp; English.  
 XX  
 CC AA298109 to AA298242 encode AAY87224 to AAY87357 which represent the  
 CC human signal peptide-containing proteins HSP-1 to HSP-134. HSPs have  
 CC anticancer, anti-inflammatory, antimicrobial, neurotrophic, hepatotropic,  
 CC neuroprotective, cardiovascular and antiasthmatic activities, and can  
 CC be used in gene therapy. HSPs can be used to treat or prevent disorders  
 CC associated with decreased activity or function of HSP. Antagonists of  
 CC HSP are used to treat or prevent disorders associated with increased  
 CC activity or function of HSP. Such diseases include cell proliferation  
 CC (including cancer), inflammation, cardiovascular, neurological,  
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,  
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,  
 CC asthma, Crohn's disease, microbial or other infections, congestive or  
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's

CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSP  
 CC nucleic acids can be used for the recombinant production of HSP, for  
 CC detecting HSP in standard hybridisation and amplification assays (for  
 CC diagnosis and monitoring), in gene therapy, as antisense,  
 CC triplex-forming or ribozyme therapeutics, for detecting related sequences  
 CC or genetic variations, and for chromosomal mapping. HSP are also used to  
 CC raise specific antibodies (Ab) and to screen for agonists and  
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or  
 CC monitor, HSP-related diseases (in usual immunoassays), as therapeutic  
 CC antagonists, in competitive drug screens, and for purification of HSP  
 CC from natural sources.  
 XX  
 SQ Sequence 383 AA:  
 Query Match 100.0%; Score 2080; DB 21; Length 383;  
 Best Local Similarity 100.0%; Pred. No. 8e-147;  
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 M A G I P G L L F L F L I C A V G S P Y S A P M K P T W P A Y R L P V L P Q S T L N L A K P D F G A E A K L E 60  
 Db 1 m a g i p g l l f l l f l l l c a v g s p y s a p m k p t w p a y r l p v l p q s t l n l a k p d f g a e a k l e 60  
 QY 61 V S S S C G P O C H K G T P L T Y E A K Q Y L S Y E T L Y A N G S R T E T O V G I I L S S G D C A Q H R D S G S 120  
 Db 61 v s s s c g p o c h k g t p l t y e a k q y l s y e t l y a n g s r t e t o v g i i l s s g d q a q h r d s g s 120  
 QY 121 S G S R R K R Q I Y G D S F S F I G K D F L N P F S T S V K T S G C T G L V A E K V L T A A R C I H D G 180  
 Db 121 s g s r r k r q i y g d s f s i f g k d f l n p f s t s v k t s g c t g l v a e k v l t a a r c i h d g 180  
 QY 181 K T Y V K G T Q L R V G F L K P R F K D G R G A N D S T S A M P E O M K F O W I R K T H V P K M I G N A N D 240  
 Db 181 k t y v k g t q l r v g f l k p r f k d g r g a n d s t s a m p e m k f o w i r k t h p k y g i l y n a n d 240  
 QY 241 I G M D Y O A L L E I K K P H K R K M I G V S P A K O L P G R I H F S G D N D R G N L Y V R F C D V K D E 300  
 Db 241 i g m d y o a l l e i k k p h k r k m i g v s p a k q l p g r i h s y g n d n p g n l y v r f c v k d e 300  
 QY 301 T Y D L L Y O C C D A O P G A S G S G Y Y V M M K R Q O O K W E R K I I G I F S G H W D M N G S P D E F N V A V R 360  
 Db 301 t y d l l y g q c d a g p a s g s g y y v m m k r g q k w e r k i i g i f s g h w d m n g s p d e f n v a v r 360  
 QY 361 I T P L K Y A Q I C Y W I K N G N L D C R E G 383  
 Db 361 i t p l k y a q i c y w i k n g n l d c r e g 383  
 RESULT 8  
 ID AAY53627 standard; Protein; 383 AA.  
 AC AAY53627;  
 DT 22-FEB-2000 (first entry)  
 DE A bone marrow secreted protein designated BMS192.  
 XX  
 KW Bone marrow secreted protein; bone marrow stromal cell; cytokine;  
 KW cell proliferation; cell differentiation; hematopoiesis; anaemia;  
 KW myeloid cell deficiency; lymphoid cell deficiency; myeloid cell;  
 KW erythroid progenitor cell; colony stimulating factor; granulocyte;  
 KW monocyte; macrophage; myelo-suppression; megakaryocyte; platelet;  
 KW platelet disorder; thrombocytopenia; bone differentiation;  
 KW stem cell disorder; aplastic anaemia; bone growth; cartilage;  
 KW paroxysmal nocturnal hemoglobinuria; bone growth; cartilage;  
 KW ligament; nerve; wound healing; tissue repair; burn; incision; ulcer;  
 KW bone fracture; cartilage damage; artificial joint.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..19





XX Human Zsig13 variant #2, SEQ ID NO:15.  
 DE  
 XX  
 KM Human Zsig13: serine protease; chromosome 11q22.1; elastase homologue;  
 KM glutamyl endopeptidase homologue; factor X homologue; trypsin homologue;  
 KM trypsinogen homologue; mast cell protease homologue;  
 KM collagenase homologue; protein degradation; food processing; brewing;  
 KM alcohol production; laundry detergent component.  
 OS Homo sapiens.  
 PN US6153420-A.  
 XX  
 XX 28-NOV-2000.  
 PD  
 XX  
 PF 04-MAY-1998; 98US-0072384.  
 XX  
 XX 24-APR-1997; 97US-0044185.  
 XX 17-APR-1998; 98US-0062142.  
 XX  
 PA (ZYMO ) ZYMOGENETICS INC.  
 XX  
 PI Sheppard PO;  
 XX  
 DR WPI: 2001-060090/07.  
 DR N-PSDB: AAC91783.  
 XX  
 PT New isolated serine protease (designated Zsig13), useful in industrial  
 PT processes to degrade unwanted proteins or alter the characteristics of  
 PT protein-containing composition, as well as in industrial applications  
 PT (e.g. brewing).  
 PS  
 PS Claim 1: Column 35-38; 26pp; English.  
 XX  
 CC The invention relates to human Zsig13 proteins (AAB48972-B48974), and  
 CC to DNA encoding them (AAC91782-C91784). The invention also relates to  
 CC expression vectors and host cells comprising a human Zsig13 DNA, and the  
 CC recombinant production of a human Zsig13 protein or its precursor.  
 CC Zsig13 is a serine protease, and has significant homology to Bacillus  
 CC licheniformis glutamyl endopeptidase, human clotting factor X, human  
 CC elastase, rat mast cell protease, Streptomyces griseus trypsin, bovine  
 CC trypsinogen, and Hypoderma lineatum collagenase. The gene encoding human  
 CC Zsig13 is located on chromosome 11q22.1. Zsig13 is useful in industrial  
 CC processes to degrade unwanted proteins or alter the characteristics of  
 CC protein-containing compositions. It may also be used in industrial  
 CC applications in which proteases are utilised, including food processing,  
 CC brewing and alcohol production, and as a component of a laundry  
 CC detergent. The present sequence represents a human Zsig13 variant.  
 XX  
 XX Sequence 392 AA:  
 SQ  
 Query Match 100.0%; Score 2080; DB 22; Length 392;  
 Best Local Similarity 100.0%; Pred. No. 8.3e-147;  
 Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAGIGLLEFLLEFFLLCAVGOVSPYSAPMKPTPAYRLPVLLPOSTLNLAKPFGAEAKLE 60  
 DB 1 maglpgllflffllcavgyvspyapkwkpwpayrlpvlpqstlnlakpfgaeakle 60  
 QY 61 VSSSGPQCHGKPTLPTEBAKQYISTETLVANGSRTEOVGITYLLSSGGCAQHRDSCS 120  
 DB 61 vsssgpqchgytpripetyeakqyisytelyangsrteqyiyllssgdaqhrds 120  
 QY 121 SGKSRKQIYGYSRFSIFGDFLLNPFSTSVKLTSGCGTLYAEKHVLTAAACHIDG 180  
 DB 121 sgkstrkqiygydsrfsifgkdfllnfpstsvxistgcvglyaeakhvltaaachidg 180  
 QY 181 KTYVGTOKLRYGLVLPKFKDGRGANDSTAMPQMKFOWIRVKTVPKGMINGAND 240  
 DB 181 ktyvgtgklrvglfklpfrfkdggrgandstampeqmkfowlrvtlvpykglkngand 240  
 QY 241 IGMDDVALLLELKRHKRKKPKMKIGVSPPAKQLPGRINHFSGIDNDRPGNLVYRFCDVADE 300

DB 241 Igmddyalllelkpkhkrkrmkiyvspakqipggrlhfsgydnrdpnlvyrfdvdx 300  
 QY 301 TYDLLYQOCDAOPGASGSQVYRMKROOKERKITIGFSCHOWNMNGSPQDENVAVR 360  
 DB 301 tydlllyqcdagpagsqgyvrmkrqgkwerkitigfshqwmngspqdenva 360  
 QY 361 ITPLEKRYAOICWIKGNVLDCCREG 383  
 DB 361 itplkyaqicwyikgnvldccreg 383  
 RESULT 12  
 AAB88336  
 ID AAB88336 standard; Protein: 383 AA.  
 XX  
 XX AAB88336;  
 AC  
 XX 23-MAY-2001 (first entry)  
 DT  
 XX  
 DE Human membrane or secretory protein clone PSEC0048.  
 XX  
 KM Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KM rheumatoid arthritis; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 PM EPI067182-A2.  
 PD 10-JAN-2001.  
 PF 07-JUL-2000; 2000EP-0114090.  
 XX  
 PR 08-JUL-1999; 99JP-0194179.  
 PR 11-JAN-2000; 2000JP-0118775.  
 PR 02-MAY-2000; 2000JP-0183766.  
 XX  
 PA (HELI-) HELIX RES INST.  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 DR WPI: 2001-093989/11.  
 DR N-PSDB: AAF93763.  
 XX  
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development -  
 PS  
 PS Claim 1: SEQ ID 40; 609pp + CD ROM; English.  
 XX  
 XX This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by  
 CC AAB88317 - AAB88419. Included in the invention are primers  
 CC AAF93917 - AAF94295 and AAF62232 - AAF62235 which are used to isolate the  
 CC cDNA sequences of the invention. The invention also includes methods for  
 CC the production of antibodies directed against the proteins, and cDNA  
 CC sequences, which can be used in vaccines. The polynucleotide sequences  
 CC can be used in gene therapy. The polynucleotide sequences and the  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with inappropriate secretory  
 CC protein/membrane protein expression. The nucleic acids and complementary  
 CC sequences may also be used as DNA probes in diagnostic assays  
 CC (e.g. polymerase chain reactions (PCR)) to detect and quantitate the  
 CC presence of similar nucleic acid sequences in samples. They may also be  
 CC used to study the expression and function of secretory proteins/membrane  
 CC polypeptides and their role in metabolism. The polypeptides may be used  
 CC as antigens in the production of antibodies against them and in assays to  
 CC identify modulators (agonists and antagonists) of expression and  
 CC activity. The antibodies and antagonists may also be used as therapeutic  
 CC agents to down regulate expression and activity. The antibodies may also  
 CC be used as diagnostic agents for detecting the presence of the  
 CC polypeptides in samples (e.g. by enzyme linked immunosorbent assay  
 CC (ELISA). Examples of diseases which may be treated include rheumatoid  
 CC arthritis and diabetes.









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 11:39:10 ; Search time 12.21 Seconds  
(without alignments)  
631.891 Million cell updates/sec

Title: US-09-554-933-3  
Perfect score: 2080  
Sequence: 1 MAGIPULFLFLCAVQG.....LKYAQCWIKGNLDCREG 383

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Archived: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfilest.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2080	100.0	383	4	US-09-072-384-18
2	2080	100.0	392	4	US-09-072-384-15
3	2044	98.3	392	4	US-09-072-384-2
4	177	8.3	222	1	US-08-090-048-1
5	177	8.5	222	2	US-08-292-550-1
6	177	8.5	222	2	US-07-927-661A-1
7	122	5.9	256	3	US-08-906-769-89
8	122	5.9	256	4	US-08-906-616-89
9	122	5.9	256	4	US-08-817-795-89
10	122	5.9	256	4	US-08-639-075A-89
11	122	5.9	256	4	US-09-012-431-89
12	122	5.9	256	4	US-09-032-215-32
13	122	5.9	256	4	US-09-012-692-89
14	122	5.9	256	4	US-08-906-613-89
15	122	5.9	256	5	PT-US95-14442A-89
16	117	5.6	437	1	US-08-487-037-2
17	115.5	5.5	241	4	US-08-944-483-59
18	114	5.5	228	4	US-08-944-483-44
19	114	5.5	253	6	5223425-8
20	113	5.4	238	6	5223425-5
21	113	5.4	250	6	5223425-4
22	112.5	5.4	223	1	US-08-278-091-13
23	112.5	5.4	223	1	US-08-483-859-13
24	112.5	5.4	223	1	US-08-472-173-13
25	112.5	5.4	223	2	US-08-487-167-13
26	112.5	5.4	223	2	US-08-482-816-13
27	112.5	5.4	223	2	US-08-296-149-13

28	112.5	5.4	223	2	US-08-801-499-13	Sequence 13, Appl
29	112.5	5.4	223	3	US-08-615-271-13	Sequence 13, Appl
30	112.5	5.4	223	3	US-09-074-660-13	Sequence 13, Appl
31	112.5	5.4	223	3	US-09-074-659-13	Sequence 13, Appl
32	112.5	5.4	223	4	US-09-106-468-13	Sequence 13, Appl
33	112.5	5.4	223	4	US-09-106-466A-13	Sequence 13, Appl
34	112.5	5.4	223	4	US-09-106-467-13	Sequence 13, Appl
35	111.5	5.4	241	1	US-08-330-978-4	Sequence 4, Appl
36	111.5	5.4	241	1	US-08-474-042-4	Sequence 4, Appl
37	111.5	5.4	241	1	US-08-484-558-4	Sequence 4, Appl
38	111.5	5.4	241	1	US-08-774-592-4	Sequence 4, Appl
39	111.5	5.4	254	1	US-08-330-978-3	Sequence 3, Appl
40	111.5	5.4	254	1	US-08-474-042-3	Sequence 3, Appl
41	111.5	5.4	254	1	US-08-484-558-3	Sequence 3, Appl
42	111.5	5.4	254	1	US-08-774-592-3	Sequence 3, Appl
43	111.5	5.4	306	1	US-08-330-978-1	Sequence 1, Appl
44	111.5	5.4	306	1	US-08-474-042-1	Sequence 1, Appl
45	111.5	5.4	306	1	US-08-484-558-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-072-384-18  
Sequence 18, Application US/09072384  
Patent No. 6153420  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES  
TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Zymogenetics, Inc.  
STREET: 1201 Eastlake Avenue East  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98102  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072.384  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, Gary E. 31,648  
REGISTRATION NUMBER: 97-16C1  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6673  
TELEFAX: 206-442-6678  
TELEX:  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 383 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Signal Sequence  
LOCATION: 1..19  
OTHER INFORMATION:  
US-09-072-384-18

```

Query Match: 100.0%; Score 2080; DB 4; Length 383;
Best Local Similarity 100.0%; Pred. No. 4, 6e-223;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps
QY 1 MAGIPGLFLFLFLLCAVGVSPYSAPWKPPTWAPYRLPVYLPOSTLNLAKPDEGAEAKLE 60
Db 1 MAGIPGLFLFLFLLCAVGVSPYSAPWKPPTWAPYRLPVYLPOSTLNLAKPDEGAEAKLE 60
QY 61 VSSSCGQCHKGTPLPYEEAKOYLSTETLYANGSKRETQYGIYLLSSGDGAQHRSGS 120
Db 61 VSSSCGQCHKGTPLPYEEAKOYLSTETLYANGSKRETQYGIYLLSSGDGAQHRSGS 120
QY 121 SGRRRRRKOIYGVSRFSIFGKDELNLVNPSTSVKLSITGCGTIVAKHVLTAHCHDG 180
Db 121 SGRRRRRKOIYGVSRFSIFGKDELNLVNPSTSVKLSITGCGTIVAKHVLTAHCHDG 180
QY 181 KTYVKGTOKLRVGLKRFKFDGKGANDSTSAMPEQMKFQWIRVYKRTHVKGWIKGNAND 240
Db 181 KTYVKGTOKLRVGLKRFKFDGKGANDSTSAMPEQMKFQWIRVYKRTHVKGWIKGNAND 240
QY 241 IGMIDYVALLLEKRPKHKRKPKITGVSPAKOLPGGRILHPSGYNDRCGNLYVRCYKDE 300
Db 241 IGMIDYVALLLEKRPKHKRKPKITGVSPAKOLPGGRILHPSGYNDRCGNLYVRCYKDE 300
QY 301 TYDILYQOCDAOPASGSGVYVFMKRRQOQKWEKKIIGIFSGHQMVMNMGSPDFENAVR 360
Db 301 TYDILYQOCDAOPASGSGVYVFMKRRQOQKWEKKIIGIFSGHQMVMNMGSPDFENAVR 360
QY 361 ITPLKVAOICYWIKGNVLDRCRG 383
Db 361 ITPLKVAOICYWIKGNVLDRCRG 383

RESULT 2
US-09-072-384-15
Sequence 15, Application US/09072384
Patent No. 6153420
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
TITLE OF INVENTION: SERINE PROPEASE POLYPEPTIDES
TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zymogenetics, Inc.
STREET: 1201 Eastlake Avenue East
City: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTED for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072.384
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31,648
REFERENCE/DOCKET NUMBER: 97-16C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-442-6673
TELEFAX: 206-442-6678
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids

```

```

: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE: internal
: FEATURE:
: NAME/KEY: Signal Sequence
: LOCATION: 1...19
: OTHER INFORMATION:
US-09-072-384-15

Query Match      100.0%; Score 2080; DB 4; Length 392;
Best Local Similarity 100.0%; Pred. No. 4.8e-223;
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 MAGTIGLLEFLLELCANQGVSPYSAPMKPTWPAYRLPYVLPOSTILNLAKPDPEGAEKLE 60
DB 1 MAGTIGLLEFLLELCANQGVSPYSAPMKPTWPAYRLPYVLPOSTILNLAKPDPEGAEKLE 60
QY 61 VSSSGPOCHKPTLPTEAEAKQVLSYETLVANGSKTETQVGIYILSSSGDAQHRDSSG 120
DB 61 VSSSGPOCHKPTLPTEAEAKQVLSYETLVANGSKTETQVGIYILSSSGDAQHRDSSG 120
QY 121 SKSKRRKQIYGYDSRFSIFGKDFLLNPFSTSVKLTSCCTIVAERKVLVAARHCIHG 180
DB 121 SKSKRRKQIYGYDSRFSIFGKDFLLNPFSTSVKLTSCCTIVAERKVLVAARHCIHG 180
QY 181 KTYVVGTOQLRVLGFLPKPKEDGGRGANDSTAMPEDMKFQWIRKTRHVPKGVKIGNAND 240
DB 181 KTYVVGTOQLRVLGFLPKPKEDGGRGANDSTAMPEDMKFQWIRKTRHVPKGVKIGNAND 240
QY 241 TGMDDYALLETKKPKKKRPMKIGVSPPAKQLPGGRTHFSGYDNDPRLVYRFGCDYKDE 300
DB 241 TGMDDYALLETKKPKKKRPMKIGVSPPAKQLPGGRTHFSGYDNDPRLVYRFGCDYKDE 300
QY 301 TYDLIXYQOCDAOPGASGSGVYVRMMKROQOKKERKLTIGIFSGHOWDMANGSPQDFNVAVR 360
DB 301 TYDLIXYQOCDAOPGASGSGVYVRMMKROQOKKERKLTIGIFSGHOWDMANGSPQDFNVAVR 360

QY 361 ITPLXYAQICYWIKNYLDCCREG 383
DB 361 ITPLXYAQICYWIKNYLDCCREG 383

RESULT 3
US-09-072-384-2
: Sequence 2, Application US/09072384
: Patent No. 6153420
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: TITLE OF INVENTION: SERINE PROTEASE POLYPEPTIDES.
: TITLE OF INVENTION: AND MATERIALS AND METHODS FOR MAKING THEM
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Zymogenetics, Inc.
: STREET: 1201 Eastlake Avenue East
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98102
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0.
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/072,384
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:

```

ATTORNEY/AGENT INFORMATION:  
NAME: Parker, Gary E  
REGISTRATION NUMBER: 31,648  
REFERENCE/DOCKET NUMBER: 97-16C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-442-6673  
TELEFAX: 206-442-6678  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 392 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: Signal Sequence  
LOCATION: 1...19  
OTHER INFORMATION:  
US-09-072-384-2

Query Match 98.3%; Score 2044; DB 4; Length 392;

Best Local Similarity 98.4%; Pred. No. 4,9e-219;

Matches 377; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAGIPGLFLFLFLLCAVGVSPYSAPMKPTMPAYRLPVYLPOSTLNLAKPDGCAEAKLE 60  
DB 1 MAGIPGLFLFLFLLCAVGVSPYSAPMKPTMPAYRLPVYLPOSTLNLAKPDGCAEAKLE 60  
QY 61 VSSCGPCHGKGTPLPYTEAKOYLSELYANGSRTETOVGIYLLSSGDAOHRDGS 120  
DB 61 VSSCGPCHGKGTPLPYTEAKOYLSELYANGSRTETOVGIYLLSSGDAOHRDGS 120  
QY 121 SGSKRRKOIYGYSRSEIFPKDFLNPFTSVKLSGCTGLVAEKHVLTAACHIDG 180  
DB 121 SGSKRRKOIYGYSRSEIFPKDFLNPFTSVKLSGCTGLVAEKHVLTAACHIDG 180  
QY 121 KTYVGTOKLAVGLKPKFKGGRGANDSTAMPDOMFQWIRVKTVPKWKIGKGNAND 240  
DB 121 KTYVGTOKLAVGLKPKFKGGRGANDSTAMPDOMFQWIRVKTVPKWKIGKGNAND 240  
QY 181 KTYVGTOKLAVGLKPKFKGGRGANDSTAMPDOMFQWIRVKTVPKWKIGKGNAND 240  
DB 181 KTYVGTOKLAVGLKPKFKGGRGANDSTAMPDOMFQWIRVKTVPKWKIGKGNAND 240  
QY 241 IGMIDYVALLLEKRPKRRKFKIGVSPAKOLPGRIHFSGYDNDPGLNLYRFGDVKDE 300  
DB 241 IGMIDYVALLLEKRPKRRKFKIGVSPAKOLPGRIHFSGYDNDPGLNLYRFGDVKDE 300  
QY 301 TYDILYQCCDAOPGASGVYVRRMKRQOKWERKIIIGIFSGHGVDMNGSPDFENAVR 360  
DB 301 TYDILYQCCDAOPGASGVYVRRMKRQOKWERKIIIGIFSGHGVDMNGSPDFENAVR 360  
QY 361 ITPLKVAOICWIKGNLYDCREG 383  
DB 361 ITPLKVAOICWIKGNLYDCREG 383

RESULT 4  
US-08-090-048-1

Sequence 1, Application US/08090048

Patent No. 5523237

GENERAL INFORMATION:

APPLICANT: Budtz, Peter

APPLICANT: Nielsen, Per M.

TITLE OF INVENTION: PROTEIN PREPARATIONS

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5523237 disk of No. 5523237th America, Inc.

STREET: 405 Lexington Avenue, 62nd Floor

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/090,048  
FILING DATE: 16-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DK 199/91  
FILING DATE: 06-FEB-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK92/00036  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Agtis, Cheryl H.  
REGISTRATION NUMBER: 34,086  
REFERENCE/DOCKET NUMBER: 3396, 214-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-867-0123  
TELEFAX: 212-867-0298  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 222 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-090-048-1

Query Match 8.5%; Score 177; DB 1; Length 222;

Best Local Similarity 25.5%; Pred. No. 1.6e-11;

Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 9;

QY 148 YPFTSVLST---GCTTVALEKHVLTAAACHIDGKTYKGTOKLAVGLKPKFRDGG 203  
DB 17 YPRAIVHISISSGCTGWMGPTVATAGHCIXDTSSGFAGRATVSPG----- 66  
QY 204 RGANDSTAMPDOMFQWIRVKTVPKWKIGKGNANDIGMDYVALLLEKRPKRRKFKI 263  
DB 67 ---RNGTS-----YPGVSXSTRFIPSGWRSQNTN-----YDGAITLSPDIGNTYIF 113  
QY 264 GVSPPAKOLPGRIHFSGYDNDPGLNLYRFGDVKDE--VKDETYDILYQCCDAOPGASGVY 321  
DB 114 GYSYTSLSLVGTIVYISGPDGKTAGTQWQSGPIAISEYTKLOYAM-DYGGGSGSPVF 172  
QY 322 VRMKRQOKWERKII---GIFSGHGVDMNGSPDFENAVRITPLKVAOICW 372  
DB 173 EQSSRTNCSGPCSLAVHTNGVYG-----SSYNRGTRITKEVFNILTNW 217

RESULT 5  
US-08-292-550-1

Sequence 1, Application US/08292550

Patent No. 5863573

GENERAL INFORMATION:

APPLICANT: Dammann, Claus

APPLICANT: Budtz, Peter

TITLE OF INVENTION: A PROCESS FOR PRODUCING CHEESE

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5863573 disk of No. 5863573th America, Inc.

STREET: 405 Lexington Avenue, Suite 6400

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10174-6401

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/292,550  
FILING DATE: 18 AUG 1994  
CLASSIFICATION: 426  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/927,624  
FILING DATE: 31 AUG 1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DK91/0068  
FILING DATE: 8 MAR 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Gregg, Valeta A.  
REGISTRATION NUMBER: 35,127  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 867 0123  
TELEFAX: 212 867 0298  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 222 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-292-550-1

Query Match	8.5%	Score 177;	DB 2;	length 222;
Best Local Similarity	25.5%	Pred. No. 1.6e-11;		
Matches 60;	Conservative 27;	Mismatches 104;	Indels 44;	Gaps 9

[illegible]

RESULT 6  
 US-07-927-661A-1  
 Sequence 1, Application US/07927661A  
 Patent No. 5866357  
 GENERAL INFORMATION:  
 APPLICANT: Dambmann, Claus  
 APPLICANT: Mortensen, Steen  
 APPLICANT: Budtz, Peter  
 APPLICANT: Eriksen, Sverdr  
 TITLE OF INVENTION: Protein Hydrolsates  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: No. 5866357o No. 5866357th America  
 STREET: 405 Lexington Avenue  
 CITY: New York  
 STATE: NY  
 COUNTRY: USA  
 ZIP: 10174  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSD for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/927,661A

FILING DATE: 31-AUG-1992  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Agiris, Cheryl H  
 REGISTRATION NUMBER: 34,086  
 REFERENCE/DOCKET NUMBER: 3396.204-US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-867-0123  
 TELEFAX: 212-878-9655  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 222 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: NO. 5866357e  
 US-07-927-661A-1

Query Match	8.5%	Score 177;	DB 2;	length 222;
Best Local Similarly	25.5%	Pred. No. 1.6e-11;		
Matches 60;	Conservative 27;	Mismatches 104;	Indels 44;	Gaps 9;

```

QY 148 YPSTSVKLTST---GCGTLVAEHVLTAAHCCHDKPT-VYKGQKRVSELPKRFKRDG 203
Db 17 YPRAIHHISSIGSGGMWIGKPTVAATAGCICYDTSSGSAFATVSPG-----66
QY 204 RGANDSTSAMPEQMKFQWIRVYKRTFHVYKGMWIKGNANDIGMDYDALLLEKPKRHKEMKI 263
Db 67 ---KNGTS-----YPGSVKSTRFFILSGMWSGNTN-----YDGAILESEPIGNFTGYE 113
QY 264 GVSPPAKOLPEGRIHFSGYDNDPRGNTLYRFCD--VDEFTYDLIYQCCDAPGASGSGVY 321
Db 114 GYSTTSSLTGTTVITISGYPGDKRAGTOWOHSGLPAISSETYKLOYAN-DYVGQSGSGSPVF 172
QY 322 VRMKRQOQKNERKIKI---GIEFGHQWVDANGSPDFENAVARTITPKYAQICWY 372
Db 173 EOSSRRTNCGSPCLSAVHTNGVYGG-----SSYNGRTRITREVFNLNTLW 217

```

RESULT 7  
 US-08-906-769-89  
 Sequence 89, Application US/08906769  
 Patent No. 6077687  
 GENERAL INFORMATION:  
 APPLICANT: Griveau, Robert B.  
 APPLICANT: Rushlow, Keith E.  
 APPLICANT: Wu Hunter, Shirley  
 APPLICANT: Frank, Glenn R.  
 APPLICANT: Stiegler, Gary  
 APPLICANT: Gaines, Patrick J.  
 APPLICANT: Silver, Gary  
 TITLE OF INVENTION: RNA PROTEASE PROTEINS, NUCLEIC ACID  
 TITLE OF INVENTION: MOLECULES AND USES THEREOF  
 NUMBER OF SEQUENCES: 190  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sheridan Ross & McIntosh  
 STREET: 1700 Lincoln Street, Suite 3500  
 CITY: Denver  
 STATE: Colorado  
 COUNTRY: USA  
 ZIP: 80203  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/906,769  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:



REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 256 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-817-795-89

Query Match 5.9%; Score 122; DB 4; Length 256;  
Best Local Similarity 24.1%; Pred. No. 2.8e-05;  
Matches 49; Conservative 30; Mismatches 68; Indels 56; Gaps 10;

QY 101 VGIYIILSS--SGGQAHNRDSSGSKRRKQIYGYDSRFISFGKDFILNYPFSTSVKLTST 158  
DB 8 VGLSAVSSYKIKDGLDRIVGGODADIAR--YGYASLOVFNHF----- 50  
159 GCTGLVAEKHVLTAACHIDGKTYVGTOKLRVGLKPKFGDGRGANDSTSAPEQMK 218  
DB 51 -CGASILNNWYITAAHCIDYEFTY-----SVRVG---TSFQ-GRGSVHPXAQIILKHPA 100  
QY 219 FQWIRKRTHPKGMKGNANDIGMDYDYLLELKKP---HKRKFMKIGVSPPAKQLPFG 275  
DB 101 Y-----GNVTDI--DMEKALIKVRPRLNRTVTVTKLTIDVGKDMPSG 142  
QY 276 RIHF-----SGYDNDRPGNLVY 292  
DB 143 ELATVTGKGNLGEDDEDDPEQLQY 165

RESULT 10  
US-08-639-075A-89  
Sequence 89, Application US/08639075A  
Patent No. 6150125  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/639,075A  
FILING DATE: 24-APR-1996  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 256 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-639-075A-89

Query Match 5.9%; Score 122; DB 4; Length 256;  
Best Local Similarity 24.1%; Pred. No. 2.8e-05;  
Matches 49; Conservative 30; Mismatches 68; Indels 56; Gaps 10;

QY 101 VGIYIILSS--SGGQAHNRDSSGSKRRKQIYGYDSRFISFGKDFILNYPFSTSVKLTST 158  
DB 8 VGLSAVSSYKIKDGLDRIVGGODADIAR--YGYASLOVFNHF----- 50  
159 GCTGLVAEKHVLTAACHIDGKTYVGTOKLRVGLKPKFGDGRGANDSTSAPEQMK 218  
DB 51 -CGASILNNWYITAAHCIDYEFTY-----SVRVG---TSFQ-GRGSVHPXAQIILKHPA 100  
QY 219 FQWIRKRTHPKGMKGNANDIGMDYDYLLELKKP---HKRKFMKIGVSPPAKQLPFG 275  
DB 101 Y-----GNVTDI--DMEKALIKVRPRLNRTVTVTKLTIDVGKDMPSG 142  
QY 276 RIHF-----SGYDNDRPGNLVY 292  
DB 143 ELATVTGKGNLGEDDEDDPEQLQY 165

RESULT 11  
US-09-012-431-89  
Sequence 89, Application US/09012431  
Patent No. 6180383  
GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Wu Hunter, Shirley  
APPLICANT: Frank, Glenn R.  
APPLICANT: Stiegler, Gary  
APPLICANT: Gaines, Patrick J.  
APPLICANT: Silver, Gary  
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/012,431  
FILING DATE: 23-Jan-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:



Query Match	5.9%;	Score 122;	DB 4;	Length 256;
Best Local Similarity	24.1%;	Pred. NO. 2.8e-05;		
*Matches	49;	Conservative	30;	Mismatches 68;
				Indels 56;
				Gaps 10;

[illegible]

RESULT 14  
US-08-906-613-89  
; Sequence 89, Application US/08906613

```

1  GENERAL INFORMATION:
2  APPLICANT: Grieve, Robert B.
3  APPLICANT: Rushlow, Keith E.
4  APPLICANT: Wu Hunter, Shirley
5  APPLICANT: Frank, Glenn R.
6  APPLICANT: Stiegler, Gary
7  APPLICANT: Gaines, Patrick J.
8  APPLICANT: Silver, Gary
9  TITLE OF INVENTION: FLA PROTEASE PROTEINS, NUCLEIC ACID
10 TITLE OF INVENTION: MOLECULES AND USES THEREOF
11 NUMBER OF SEQUENCES: 190

```

ADDRESS: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: patentin Release #1.0, Version #1.30

```

APPLICATION NUMBER: US/08/906,613  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/639,075  
FILING DATE: 24-APR-1996  
ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-25-C2  
TELECOMMUNICATION INFORMATION:

TELEFAX: (303) 863-0223  
; INFORMATION FOR SEQ ID NO: 89:

```

; LENGTH: 256 amino acids
; TYPE: amino acid
;

```

MOLECULE TYPE: protein  
US-08-906-613-89

Query Match	5.9%;	Score 122;	DB 4;	Length 256;
Best Local Similarity	24.1%;	Pred. No. 2.8e-05;		
Matches 49;	Conservative 30;	Mismatches 68;	Indels 56;	Gaps 10;

```

QY 101 VGIYIISS--SGOAGHNRDSSGSSKSRKROIQIYGDYRFSIEFKDEFLINPEFSTVSLK 158
      | | : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 VGLSAVSSYKIKDGLDRIGYGGDADIAK---YGOASLQVFNHF-----50

QY 159 GCTSTLAEKHVLTPAAHCIDHGKTYVKTOLRYGFLKPKFKDGCAGANDSTAMPEOMK 218
      | : : : : | | | | | | | | | | | | | | | | | | | | | | | |
Db 51 -CGASILNNWIVTAACHIDEFY-----SVRG--TSFO-GRGSGVPHXOILIKHPA 100

QY 219 FQWIRKRVKTHVPKCKWIGNANDIGMDYDIALLELKP--HKRKEMKIGVSPPAKOLPG 275
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 101 Y-----GNVTDI--DNEXALIKVRPEPLNNRVRVTKLIDYGDKMPG 142

QY 276 RIHF-----SGYNDRPRGNLV 292
      | : : : : | | | | | | | | | | | | | | | | | | | | | |
Db 143 ELATVTCGNLGEDDEDEPEQLOY 165
      | : : : : | | | | | | | | | | | | | | | | | | | | | |

```

PCT-US95-14442A-89  
; Sequence 89, Application PC/TUS9514442A

GENERAL INFORMATION:  
APPLICANT: Grieve, Robert B.  
APPLICANT: Rushlow, Keith E.  
APPLICANT: Hunter, Shirley Wu  
APPLICANT: Frank, Glenn R.  
APPLICANT: Heath, Andrew W.  
APPLICANT: Yamaka, Miles Yamanaka  
APPLICANT: Arfsten, Ann  
APPLICANT: Dale, Beverly  
APPLICANT: Stiegler, Gary  
TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND  
TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA  
TITLE OF INVENTION: INDESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACIDIC  
TITLE OF INVENTION: MOLECULES, AND USES THEREOF  
NUMBER OF SEQUENCES: 119

ADDRESSEE: Sheridan Ross & McIntosh  
STREET: 1700 Lincoln Street, Suite 3500  
CITY: Denver  
STATE: Colorado  
COUNTRY: USA

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: PatentIn Release #1.0, Version #1.25
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: PCT/US95/14442A

```

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER:

TELEPHONE: (303) 863-9700  
TELEFAX: (303) 863-0223  
INFORMATION FOR SEQ ID NO: 89:

```

; LENGTH: 256 amino acids
; TYPE: amino acid
;

```

MOLECULE TYPE: protein  
PCT-US95-14442A-89

Query Match	5.9%;	Score 122;	DB 5;	Length 256;
Best Local Similarity	24.1%;	Pred. No. 2.8e-05;		





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 11:40:15 ; Search time 15.96 seconds

(without alignments)  
1827.998 Million cell updates/sec

Title: US-09-554-933-3

Percent score: 2080

Sequence: 1 MGRIPLFLFLCAVQ.....LKYYQICWIKNYLDCREG 383

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-68:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

#### SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	194	9.3	316	2 A45134	endopeptidase (EC
2	135.5	6.5	313	2 A35122	metalloprotease
3	120.5	5.8	482	1 EXRT	coagulation factor
4	119	5.7	269	2 A26823	pancreatic elastase
5	117	5.6	522	2 T29767	pancreatic elastase
6	115.5	5.6	269	2 B26823	pancreatic elastase
7	115.5	5.6	271	2 A25528	pancreatic elastase
8	115	5.5	258	2 S70439	pancreatic elastase
9	115	5.5	267	4 A56615	probable pancreati
10	114.5	5.5	238	1 TRWV5T	trypsin-like prote
11	114	5.5	246	1 DBHU	complement factor
12	113	5.4	266	1 ELPT1	pancreatic elastase
13	113	5.4	259	1 TRSMG	pancreatic elastase
14	112.5	5.4	273	2 HE4915	hypothetical prote
15	112.5	5.4	273	2 E85765	hypothetical prote
16	112.5	5.4	273	2 HE4915	hypothetical prote
17	111.5	5.4	488	1 EXHU	coagulation factor
18	111	5.3	761	2 JC5759	brain-specific ser
19	110	5.3	492	1 EXBO	pancreatic elastase
20	110	5.3	492	1 EXBO	coagulation factor
21	110	5.3	1582	2 T15308	hypothetical prote
22	109.5	5.3	405	2 T35117	probable secreted
23	108	5.2	583	2 A29154	complement factor
24	108	5.2	786	2 A47547	serine proteinase
25	107.5	5.2	1047	2 A55617	masquerade precurs
26	107.5	5.2	236	2 A28566	T-cell suppressor
27	107.5	5.0	686	1 A59271	Re-activator factor
28	105	5.0	274	2 S40004	trypsin-related pr
29	103.5	5.0	271	1 ELRT2	pancreatic elastase

30	103.5	5.0	416	1 KFBO	coagulation factor
31	102.5	4.9	1238	2 T34929	hypothetical prote
32	100	4.8	272	2 JC4170	trypsin-like prote
33	98.5	4.7	267	2 S40006	trypsin (EC 3.4.21
34	98	4.7	548	2 D82175	probable trypsin V
35	98	4.7	624	2 T02289	probable polygalac
36	97.5	4.7	409	2 T35118	probable secreted
37	97.5	4.7	452	1 A30351	coagulation factor
38	97.5	4.7	151579	2 A38738	coagulation factor
39	97	4.7	1019	2 A38738	coagulation factor
40	96.5	4.6	259	2 S68424	allergen Der f III
41	96.5	4.6	275	2 T46712	factor IX - rabbit
42	96.5	4.6	2145	2 JC4747	adenylate cyclase
43	95.5	4.6	430	1 A24702	serine proteinase
44	95	4.6	782	2 T32155	hypothetical prote
45	94	4.5	268	2 S68825	pancreatic elastase

#### ALIGNMENTS

##### RESULT 1

A45134

endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis

C:Species: Bacillus licheniformis

C:Date: 10-Jun-1993 #sequence.revision 18-Nov-1994 #text.change 15-Oct-1999

C:Accession: A45134; S23078

R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S

J. Biol. Chem. 267, 23782-23788, 1992

A:Title: Purification, characterization, cloning, and expression of a glutamic acid-s

A:Reference number: A45134; MUID:93054737

A:Accession: A45134

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-316 <KAK>

A:Cross-references: GB:D10060; NID:g216263; PIDN:BA00949.1; PID:g21001415; PID:g21626

A:Experimental source: ATCC 14580

A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBI:118785)

R:Syvendsen, I.; Bredam, K.

Eur. J. Biochem. 204, 165-171, 1992

A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase

A:Reference number: S23078; MUID:92135199

A:Accession: S23078

A:Status: preliminary

A:Molecule type: protein

A:Residues: 95-316 <SYE>

C:Keywords: hydrolase

Query Match 9.3%; Score 194; DB 2; Length 316;

Best Local Similarity 24.0%; Pred. No. 1.2e-08;

Matches 79; Conservative 43; Mismatches 133; Indels 64; Gaps 15;

QY	63	SSCGPOCHKGPPL---	PPYEAKQOLSTETLYANGSRTEOVGIYIISSSDGAQHRDSC	119
DB	28	AQAAPSPH--TPVSSDPSTY--	KAETSVTPDP-----NIKSDQYLSKAFDTGVVNETKE	79
QY	120	SSGSRKRRQLYGDSRSIFGKDFLN---	YPESTVSKLST---GCGTVAEKHV	170
DB	80	KAEEKSPAKAPY---	SIKSVIGSDRRTVNTTAYPRATIHSSISGSGGMIGPKTV	136
QY	171	LTAACIHDKGT--YVGTQKLRVGLKPKFKDGRGANDSTAMP	EQMKPQWIRKRTHV	229
DB	137	ATAGCITDTSSTSGFAGTATVSPG-----	-----RNGTS-----YPIGSAVSTRFI	178
QY	230	PKGWIKGANDIGMDYDALLLEKPKRKRKMGVSPPAKOLF	PGRIHNSGYNDNRPN	289
DB	179	PSGWSGNTN-----YDGAILELSEPIGNTVGYGYSVTS	SLVGTVTJISGPGDVTAG	233
QY	290	LVYRPGD--VKDETYDILLOCCDQAPGASGSGVVRMKK	RQOQWEMKII---GIFSGH	343
DB	234	TOMHSGPILAISETYKLOYAM--DTYGGSGSPVEQSS	SRKNCGPCSLAVHTNGVYGG--	291

OY 344 QWYDMNGSPQDENAVAVITPELKYAOICYW 372  
 DB 292 -----SSYNGRTRITKEVFDNLTNW 311

## RESULT 2

A:Accession: A35122  
 C:Species: Bacillus subtilis  
 C:Date: 27-Jul-1990 #sequence\_revision 27-Jul-1990 #text\_change 15-Oct-1999  
 C:Accession: A35122; 140010; A69660  
 R:Stroma, A.; Rudolph, C.F.; Rido Jr., G.A.; Sullivan, B.J.; Theriault, K.A.; Ally, D.; F  
 J. Bacteriol. 172, 1024-1029, 1990  
 A:Title: Gene encoding a novel extracellular metalloprotease in Bacillus subtilis.  
 A:Reference number: A35122; MUID:90130256  
 A:Accession: A35122  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-313 <SLD>  
 A:Cross-references: GB:L10505; NID:g143209; PIDN:AAA22604.1; PID:g143210; GB:M29036  
 R:Smith, H.; de Jong, A.; Bron, S.; Venema, G.  
 Gene 70, 351-361, 1988  
 A:Title: Characterization of signal-sequence-coding regions selected from the Bacillus s  
 Reference number: I39994; MUID:89108019  
 A:Accession: 140010  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-60, 65, 'L', 67, 'S', 69, 'AQA' <RES>  
 A:Cross-references: GB:M22916; NID:g143701; PIDN:AAA22832.1; PID:g143702  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berth  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
 A.; Ehrlich, S.D.; Emerson, P.T.; Estlin, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: A69660  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-313 <KUB>  
 A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12018.1; PID:e1182176;  
 A:Experimental source: strain 168  
 Genetics:  
 A:Gene: mpr  
 C:Keywords: hydrolase

Query Match 6.5%; Score 135.5; DB 2; Length 313;  
 Best Local Similarity 22.2%; Pred. No. 0.00991;  
 Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;  
 OY 72 GHPLEPYEAKOYLSEYELLYANGSRTEOVGIYLISSGSD--GAOHRDSSGSGKRRKRO 129  
 DB 29 GVPAAKAENPQTSVNTGKADATKNOT-----SKADQVSAPVEGTGKTSK----- 75  
 OY 130 IYGYDSRF-----SIFGKD-----FLLNPEFSVSLST-----CCTGTL 164  
 DB 76 LVGGQLELEKNITLOPSSIIIGDERTRLSSTTSFPRATVQLSKYPTTSYTGCTGL 135  
 OY 165 VAEKHVLTAAHCH-----DGKTYVKGTOKLRYGFLKPKFGRGAND 208  
 DB 136 VNNNTYVTAACHYVCSODHGMASHTTAPARNSSPYGY----- 175  
 OY 209 STSAMPEQMKFQWIRVAKRTHVPKWIKGN---ANDIGMDYVALLLELKKPKRKEKIGV 265

DB 176 -SGTFYISVK-CWTESKOTNDYGAIKNGSPGNTVGH-YCIRRTNSSP-----VGL 225  
 OY 266 SPFAKQLPGRIHFGSYDNRPGNLVYRFCDVDEYDLYOQCDAPGASGSGYVRAW 325  
 DB 226 SSSVYGFPODKRFGTMWSDPKIR-----SAETFKLFY-TTDYGCOSGSPY----- 272

OY 326 KROQKWEKIKTIFSGHGWDMNGSPQDENAVAVITPELKYAOICYW 372  
 DB 273 -RNSYDTGQFALAIHT-----NCG-SSYNGTRVNDVFNNTIOW 310

## RESULT 3

EXRPT  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Jan-1995 #sequence\_revision 07-Feb-1997 #text\_change 08-Dec-2000  
 C:Accession: S49075; J04670; P50191; P50190; 162745  
 R:Stanton, C.; Ross, R.P.; Hutson, S.; Wallin, R.  
 Thromb. Res. 80, 63-73, 1995  
 A:Title: Evidence for competition between vitamin K-dependent clotting factors for in  
 A:Reference number: A58498; MUID:96093366  
 A:Accession: S49075  
 A:Molecule type: mRNA  
 A:Residues: 1-482 <STA1>  
 A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601  
 A:Note: submitted to the EMBL Data Library, June 1994  
 A:Note: neither the complete nucleic acid sequence nor the complete translation are s  
 Gene 169, 269-273, 1996  
 R:Stanton, C.; Ross, R.P.; Hutson, S.; Wallin, R.  
 A:Title: Processing and expression of rat and human clotting factor-X-encoding cDNAs.  
 A:Reference number: J04670; MUID:96194815  
 A:Accession: J04670  
 A:Molecule type: mRNA  
 A:Residues: 1-482 <STA2>  
 A:Cross-references: EMBL:X79807; NID:g506600; PIDN:CAA56202.1; PID:g506601  
 A:Experimental source: Cos-1 cell  
 R:Enjiyoji, K.; Miyazaki, K.; Kato, H.  
 J. Biochem. 109, 890-898, 1991  
 A:Title: Characterization of rat factors X and Xa: demonstration of factor Xa in rat  
 A:Reference number: P50190; MUID:92041742  
 A:Accession: P50191  
 A:Molecule type: protein  
 A:Residues: 41-56, 'X', 60-65 <ENUJ>  
 A:Accession: P50190  
 A:Molecule type: protein  
 A:Residues: 183-186, 'X', 188-207 <ENU2>  
 A:Residues: 183-186, 'X', 188-207 <ENU2>  
 R:Murakawa, M.; Okamura, T.; Kamura, T.; Kuroiwa, M.; Harada, M.; Niho, Y.  
 Eur. J. Hematol. 52, 162-168, 1994  
 A:Title: Analysis of the partial nucleotide sequences and deduced primary structures  
 A:Reference number: I46196; MUID:94222160  
 A:Accession: 162745  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 295-383, 'G', 385-455 <MUR>  
 A:Cross-references: GB:D21215; NID:g415309; PIDN:BA04756.1; PID:g455396.  
 C:Function:  
 A:Description: catalyzes the proteolytic activation of prothrombin to thrombin in the  
 A:Pathway: blood coagulation  
 A:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol  
 C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu  
 F:1-23/Domain: signal sequence #status predicted <PRO>  
 F:24-40/Domain: propeptide #status predicted <PRO>  
 F:41-179/Product: coagulation factor X light chain #status predicted <LCH>  
 F:129-144/Domain: EGF homology <EG1>  
 F:129-144/Domain: EGF homology <EG2>  
 F:183-482/Product: coagulation factor X heavy chain #status predicted <HCH>  
 F:183-221/Domain: activation peptide #status predicted <APT>  
 F:232-482/Product: coagulation factor Xa heavy chain #status predicted <ACT>  
 F:232-480/Domain: trypsin homology <TRI>  
 F:46, 47, 54, 56, 59, 66, 69, 72, 79/Modified site: gamma-carboxyglutamic acid (Glu) #s  
 F:57-62, 90-101, 95-110, 112-121, 129-140, 136-149, 151-164, 173-340, 238-243, 259-275, 368-402







A:Residues: 1-266 <MAC>  
A:Cross-references: GB:V01234; NID:956088; PIDN:CAA24544.1; PID:956089  
R:Largman, C.  
Biochemistry 22, 3763-3770, 1983  
A:Title: Isolation and characterization of rat pancreatic elastase.  
A:Reference number: A20534; MUID:84000385  
A:Accession: A20534  
A:Molecule type: protein  
A:Residues: 17-37, 'X', 39-45 <LAR>  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; pancreas; serine proteinase; zymogen  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-26/Domain: activation peptide #status predicted <ACT>  
F:27-266/Product: elastase I #status predicted <MPT>  
F:27-259/Domain: trypsin homology <TRY>  
F:71,119,214/Active site: His, Asp, Ser #status predicted

Query Match 5.4%; Score 113; DB 1; Length 266;  
Best Local Similarity 26.7%; Pred. No. 0.056;  
Matches 39; Conservative 23; Mismatches 48; Indels 36; Gaps 7;

128 RQYGVDSRPSFGKDEFLNPESTSVKLTG-----CTGTVAEKHVLTAHCIDHDKT 182  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 26 RYVGSAFARNNSPQSLQY-----LSGSGMYHTCGTLLRRNMVMTAAHCVSQMT 78  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 183 YKQYQKLRVGFLLPKFKDGRGANDSTAMPQMKFQWIRKRVKRTYVPGWIKGANDIG 242  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 79 F-----RVVVG-----DHNLSQNDGTE-----QYVSQKIVVHPVW---NSNNVA 115  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 243 MDYDALLET-KKPKRKKEMKIGSP 267  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 116 AGYDIALRLAQSVTLNNVQLAVLP 141  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 13  
ELPG  
pancreatic elastase (EC 3.4.21.36) I precursor - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 24-Apr-1984 #sequence\_revision 30-Sep-1990 #text\_change 16-Jun-2000  
C:Accession: J50013; A26777; A10061; A00959  
R:Shirasu, Y.; Yoshida, H.; Miyayama, T.; Matsuki, S.; Tanaka, J.I.; Ikenaga, H.  
J. Biochem. 99, 1707-1712, 1986  
A:Title: Isolation and expression in *Escherichia coli* of a cDNA clone encoding porcine B  
A:Reference number: A92005; MUID:86304235  
A:Accession: J50013  
A:Molecule type: mRNA  
A:Residues: 1-266 <SHI>  
A:Cross-references: GB:X04036; GB:D00070; GB:N00070; NID:91941; PIDN:CAA27670.1; PID:919  
R:Tani, T.; Kawashima, I.; Furukawa, H.; Ohmine, T.; Takiguchi, Y.  
Biochem. 101, 591-599, 1987  
A:Title: Characterization of a silent gene for human pancreatic elastase I: structure of  
A:Reference number: A26777; MUID:87250343  
A:Accession: A26777  
A:Molecule type: mRNA  
A:Residues: 1-125, 'G', 127-183, 'L', 185-266 <PAN>  
A:Cross-references: GB:D00160; NID:9217683; PIDN:BA00118.1; PID:9217684  
R:Shotton, D.M.; Hartley, B.S.  
Nature 225, 811-816, 1970  
A:Title: Three-dimensional structure of tosyl-elastase.  
A:Reference number: A93160; MUID:70114044  
C:Contents: annotation; X-ray crystallography, 3.5 angstroms; active site  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; pancreas; serine proteinase; zymogen  
F:1-16/Domain: signal sequence #status predicted <SIG>

F:17-26/Domain: activation peptide #status predicted <APR>  
F:27-266/Product: elastase I #status experimental <MAT>  
F:27-259/Domain: trypsin homology <TRY>  
F:56-72,153-200,184-200,210-240/Disulfide bonds: #status experimental  
F:71,119,214/Active site: His, Asp, Ser #status experimental

Query Match 5.4%; Score 113; DB 1; Length 266;  
Best Local Similarity 27.3%; Pred. No. 0.056;  
Matches 36; Conservative 21; Mismatches 41; Indels 34; Gaps 6;

QY 147 NYPESTSVKLTG-----CTGTVAEKHVLTAHCIDHDKTY--VKQYKLRVGFLLPK 198  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 37 SWPQSLQYREGSSWMAHTCGTLLRRNMVMTAAHCVDRELFVVGSEHNL----- 88  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 199 FQDGRGANDSTAMPQMKFQWIRKRVKRTYVPGWIKGANDIGDYALLET-KKPK 257  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 89 -----NNDGTE-----QYGVQKIVVHPVW---NTDVAAGYDIALRLAQSVTL 131  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 258 RKFMKIGVSP 269  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 132 NSYVQLGVLPRA 143  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 14  
TRSMG  
trypsin (EC 3.4.21.4) precursor - Streptomyces griseus  
C:Species: Streptomyces griseus  
C:Date: 24-Apr-1984 #sequence\_revision 12-May-1994 #text\_change 07-May-1999  
C:Accession: J01302; A00962  
R:Kim, J.C.; Cha, S.H.; Jeong, S.T.; Oh, S.K.; Byun, S.M.  
Biochem. Biophys. Res. Commun. 181, 707-713, 1991  
A:Title: Molecular cloning and nucleotide sequence of Streptomyces griseus trypsin ge  
A:Reference number: J01302; MUID:92095977  
A:Accession: J01302  
A:Molecule type: DNA  
A:Residues: 1-259 <KIM>  
A:Cross-references: GB:M64471  
A:Experimental source: strain ATCC10137  
R:Olafson, R.W.; Jurasek, L.; Carpenter, M.R.; Smillie, L.B.  
Biochemistry 14, 1168-1177, 1975  
A:Title: Amino acid sequence of Streptomyces griseus trypsin. Cyanogen bromide fragme  
A:Reference number: A00962; MUID:75127940  
A:Accession: A00962  
A:Molecule type: protein  
A:Residues: 37-95,98-259 <OLA>  
R:Read, R.J.; James, M.N.G.  
J. Mol. Biol. 200, 523, 1988  
A:Title: Refined crystal structure of Streptomyces griseus trypsin at 1.7 angstroms r  
A:Reference number: A44574; MUID:88286735  
A:Contents: annotation; X-ray crystallography, 1.7 angstroms  
A:Note: residues 96-97 modeled as Gly-Ala  
C:Genetics:  
A:Gene: SPRT  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:1-32/Domain: signal sequence #status predicted <SIG>  
F:33-36/Domain: propeptide #status predicted <PRO>  
F:37-258/Product: trypsin #status experimental <MAT>  
F:37-252/Domain: trypsin homology <TRY>  
F:56-74,177-192,204-233/Disulfide bonds: #status experimental  
F:73,118,208/Active site: His, Asp, Ser #status experimental

Query Match 5.4%; Score 112.5; DB 1; Length 259;  
Best Local Similarity 26.0%; Pred. No. 0.06;  
Matches 39; Conservative 21; Mismatches 53; Indels 37; Gaps 5;

QY 148 YPESTSVKLTGCTGTVAEKHVLTAHCIDHDKTYVKGQKLRVGFLLPKFKDGRGAN 207  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
Db 48 FPF--VYRLSMGCGGALYADIVLTAHCV-----SSGNN 81  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
QY 208 DS---TSAMPQMKFQWIRKRVKRTYVPGWIKGANDIGMDYALLETKKPKRKKMKIG 264  
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |



Db 82 TSIRATGCVVLQSSSAVKVSTKYLQA-----PGYNGTGRKDWALIKLAQPIINQPTLKIA 136  
QY 265 VSPPAKOLPGGRIRHSGYDNDPGLVYRF 294  
Db 137 TTTAVNQ---GTFYVAGWGANREGGSOQRY 163

## RESULT 15

E85765

hypothetical protein z2592 [imported] - Escherichia coli (strain O157:H7)

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 31-Mar-2001

C:Accession: E85765

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,

Nature 409, 529-533, 2001

Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

Reference number: A85480; MUID:21074935; PMID:1120551

Accession: E85765

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-273 &lt;STO&gt;

A:Cross-references: GB:AE005174; NID:912515576; PIDN:AAG56585.1; GSPDB:GN00145; UWGP:Z25

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: z2592

Query Match 5.48; Score 112.5; DB 2; Length 273;

Best Local Similarity 23.38; Pred. No. 0.064;

Matches 62; Conservative 35; Mismatches 112; Indels 57; Gaps 13;

QY 101 VGIYLLSSGGDAQHRDSSGSKSRKROIYGYDSRFSIFGKDFLLNYPFSTSVKLTSG- 159  
Db 9 LGAISLISAEYFADKPDYAKSANDDEVSTLFGHDDRPV---NDTQSPMDAVGQLETAS 65  
QY 160 ---CTGLVAEKHVLTAACIHDKTYVKGTOKLRVGLKPKFKDGR-----GANDS 209  
Db 66 GNLCTATLTAAPNLALTAGHCL---LTPPKAKADKAVLRFSNKGMLWRYDIDHIEGRVDP 122  
QY 210 TSAMEQOKFQWIRKTHVKGWIKGNANDIGMDYALFELKPKHKKRKIKGVSP-- 267  
Db 123 T--LGRRLKAD-----GDGWLIVPPA---AAPWDFGLIVLRNPPS-----GITPLP 162  
QY 268 -----PAKOLPGGRIRHSGYDNDPGLVY--YRCDVKD--ETYDILLYOCCDAOPGA 315  
Db 163 LFEQDKALLTAALKRAAGKRYTQAGYRPEDHLDITLYSHONCEVTGMAQTSYMSHQCDTLPGD 222  
QY 316 SGSQYVVRMKRQOKWKRKIIGIFS 341  
Db 223 SGSPLML---HTDQGWQ--LIGVQS 242

Search completed: July 12, 2001, 11:43:32

Job time: 197 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 11:42:36 ; Search time 13.12 Seconds  
(without alignments)  
999,989 Million cell updates/sec

Title: US-09-554-933-3  
Perfect score: 2080  
Sequence: 1 MAGIPGLFLFLFLLCAVQ.....LKYAQICWIKGNYLDCREG 383

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 93435 segs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	194	9.3	316	1	GSEP_BACLI
2	135.5	6.5	313	1	MPR_BACSU
3	124	6.0	266	1	ELI_BOVIN
4	121.5	5.8	490	1	FA10_RABIT
5	119	5.7	269	1	EL2_PIG
6	115.5	5.6	269	1	EL2A_HUMAN
7	115.5	5.6	271	1	EL2_MOUSE
8	115	5.5	376	1	FA10_TROCA
9	114.5	5.5	238	1	TRP5_AEDAE
10	113	5.4	266	1	ELI_PIG
11	113	5.4	266	1	ELI_RAT
12	112.5	5.4	259	1	TRYP_STRGR
13	112.5	5.4	488	1	FA10_HUMAN
14	111.5	5.4	875	1	NETR_HUMAN
15	111	5.3	253	1	CEAD_HUMAN
16	111	5.3	761	1	NETR_MOUSE
17	110	5.3	492	1	EL2B_HUMAN
18	110	5.3	269	1	FA10_BOVIN
19	108	5.2	583	1	CFAL_HUMAN
20	108	5.2	786	1	STUB_DROME
21	107.5	5.2	686	1	MAS2_HUMAN
22	105	5.0	269	1	EL2_BOVIN
23	105	5.0	274	1	TRP5_ANOGA
24	104.5	5.0	268	1	TRYP_STRGA
25	103.5	5.0	271	1	EL2_RAT
26	103.5	5.0	416	1	FA9_BOVIN
27	102	4.9	902	1	ST14_MOUSE
28	101.5	4.9	281	1	TRP2_DROER
29	98.5	4.7	260	1	COG8_HYPLI
30	98.5	4.7	267	1	TRP7_ANOGA
31	97.5	4.7	452	1	FA9_CANFA
32	97	4.7	1019	1	TRC_TACTRA
33	96.5	4.6	252	1	TR11_DROME

34	96.5	4.6	259	1	DEF3_DERPA
35	96.5	4.6	275	1	FA9_RABIT
36	96.5	4.6	454	1	TMS3_HUMAN
37	96.5	4.6	2145	1	CYAA_PODAN
38	96	4.6	274	1	TRP1_ANOGA
39	96	4.6	855	1	ST14_HUMAN
40	95.5	4.6	430	1	SNK_DROME
41	95	4.6	1019	1	LFC_CARO
42	94.5	4.5	261	1	DER3_DERPT
43	93.5	4.5	227	1	TRYP_SACER
44	93.5	4.5	277	1	KIKD_HUMAN
45	93	4.5	247	1	MCT5_MOUSE

## ALIGNMENTS

```

RESULT 1
GSEP_BACLI STANDARD: PRT; 316 AA.
ID AC P80057;
DF 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE GLUTAMYL ENDOPEPTIDASE PRECURSOR (EC 3.4.21.19) (GLUTAMATE SPECIFIC
DE ENDOPEPTIDASE) (GSE).
GN BLASE.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxId=1402;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 14580;
RX MEDLINE=93054737; PubMed=1429718;
RA Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,
RA Okamoto H., Shin M., Tamaki M., Yeraoka H., Tsuzuki H., Yoshida N.;
RT "Purification, characterization, cloning, and expression of a
RT glutamic acid-specific protease from Bacillus licheniformis ATCC
RT 14580."
RT J. Biol. Chem. 267:23782-23788(1992).
RN [2]
RN SEQUENCE OF 95-316.
RN MEDLINE=92155199; PubMed=1346764;
RX Svendsen I., Bredam K.;
RA "Isolation and amino acid sequence of a glutamic acid specific
RT endopeptidase from Bacillus licheniformis."
RL Eur. J. Biochem. 204:165-171(1992).
CC -I- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE
CC CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG
CC PREFERENCE FOR GLU.
CC -I- SUBCELLULAR LOCATION: SECRETED.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE
CC V8 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL database -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10060; BAA00949.1;
CC PIR: S23078; S23078.
CC PIR: A45134; A45134.
CC MEROPS: S01.271;
CC InterPro: IPR00126;
CC InterPro: IPR001254;
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00839; V8PROTEASE.
CC PROSITE: PS00672; V8_HIS; 1.
CC PROSITE: PS00673; V8_SER; 1.

```

KW Hydrolyase; Serine protease; Signal.  
 FT SIGNAL 1 2  
 FT PROPEP 94  
 FT CHAIN 95 316  
 FT ACT\_SITE 141 141  
 FT ACT\_SITE 261 261  
 FT DISULFID 126 142  
 FT DISULFID 275 279  
 SO SEQUENCE 316 AA; 33611 MW; 96D/552CB708B09 CRC64;  
 Query Match 9.38; Score 194; DB 1; Length 316;  
 Best Local Similarity 24.08; Pred. No. 4,4e-09;  
 Matches 79; Conservative 43; Mismatches 143; Indels 64; Gaps 15;  
 QY 63 SSCGPOCHKRPL---PYEEAKOYLSEETLVANGSRTEOVGIYLLSSGDAQHDSG 119  
 DB 28 AOAAPSPH--TPVSSDPY-NAETSVYDP-----NIKSDOYGLXSAFTGTGVNETKE 79  
 QY 120 SSGSRRRKQIYGYDSRFSIPGKDFLNLN-----YPESTVYKST---GCTGLVAEKHV 170  
 DB 80 KAEKSPAKKAY---SISVIGSDDRVTNTATVPYRAIVHISSISGCTGMWIGPKTV 136  
 QY 171 LTAHCHIDGKT-YKCTQKLRVGLKRFKFGDGRGANDSTSNAPKPKQWIKRTHV 229  
 DB 137 ATAGHCIDYDSSGSPAGTATVSPG-----RNGTS-----YPIGSASTRFTI 178  
 QY 230 PRGWIKGANIDGMIDYDALLKPKHKKFMKIGVSPAPQLPGRITFSGYDNDPRGN 289  
 DB 179 PSGMWSGNTN-----YDGAIELESPIGNVYGYISYTTSSIVGTIVTISGFGDKTAG 233  
 QY 290 IYVRECD--VKDETVDLLYQCDAPGASGSGVYRMKRRQOKERKIT---GIEGSH 343  
 DB 234 TQWQSHGPIAISEYKLYQAM-DYFGQSGSGSVPEQSSSRNCSGPCSLAVHTNGVYG- 291  
 QY 344 QWVDWNGSPQDENVAVTRTPPKYAQICW 372  
 DB 292 -----SSYNGRTIRLKEVFDMITW 311  
 RESULT 2  
 MPR\_BACSU STANDARD; PRT; 313 AA.  
 ID MPR\_BACSU STANDARD; PRT; 313 AA.  
 AC P39790: 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE EXTRACELLULAR METALLOPROTEASE PRECURSOR (EC 3.4.21.-).  
 GN MPR.  
 OS Bacillus subtilis.  
 Bacteria; Firmicutes; Bacillus/Clostridium group;  
 Bacillus/Staphylococcus group; Bacillus.  
 NCBI\_TaxID=1423;  
 OX [1]  
 RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=GP241;  
 RX MEDLINE=90130256; PubMed=2105291;  
 RA Sloma A., Rudolph C.F., Ruto G.A. Jr., Sullivan B.J., Theriault K.A.,  
 Ailly D., Petro J.;  
 RT "Gene encoding a novel extracellular metalloprotease in Bacillus  
 subtilis."  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-68 FROM N.A.  
 RX MEDLINE=89108019; PubMed=3145906;  
 RA Smith H., de Jong A., Bron S., Venema G.;

RT "Characterization of signal-sequence-coding regions selected from the  
 RT Bacillus subtilis chromosome."  
 RT Gene 70:351-361(1988).  
 RN [4]  
 RN SEQUENCE OF 1-10 FROM N.A.  
 RP STRAIN=168;  
 RC MEDLINE=96118702; PubMed=7496533;  
 RX Saxild H.H., Jacobsen J.H., Nygaard P.;  
 RA "Functional analysis of the Bacillus subtilis pur gene encoding  
 RT formate-dependent glycinamide ribonucleotide transferase."  
 RT Microbiology 141:2211-2218(1995).  
 CC -1 SUBUNIT: MONOMER.  
 CC -1 SUPRACELLULAR LOCATION: SECRETED.  
 CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B; ALSO KNOWN AS THE  
 CC V8 FAMILY.  
 CC -1 CAUTION: CALLED "METALLOPROTEASE", BUT CLEARLY BELONGS TO THE S2  
 CC FAMILY OF SERINE PROTEASES.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: L10505; AAA22604.1;  
 DR EMBL: AB006424; BAA33121.1;  
 DR EMBL: M22916; AAA22832.1;  
 DR EMBL: Z99105; CAB12018.1;  
 DR PIR: A35122; A35122.  
 DR HSSP: P00756; 1SGF.  
 DR MEROPS: S01.272;  
 DR Subtilist; BG10690; mpr.  
 DR Interpro; IPR000126;  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00839; V8PROTEASE.  
 DR PROSITE; PS00672; V8\_HIS; 1.  
 DR PROSITE; PS00673; V8\_SER; 1.  
 KW Hydrolyase; Serine protease; Signal; Zymogen.  
 FT SIGNAL 1 34  
 FT PROPEP 35 93  
 FT CHAIN 94 313  
 FT ACT\_SITE 146 146  
 FT ACT\_SITE 267 267  
 FT DISULFID 131 147  
 FT CONFLICT 61 68  
 SO SEQUENCE 313 AA; 33842 MW; D41788E8D652AE94 CRC64;  
 Query Match 6.58; Score 135.5; DB 1; Length 313;  
 Best Local Similarity 22.24; Pred. No. 0.00035;  
 Matches 77; Conservative 38; Mismatches 121; Indels 111; Gaps 17;  
 QY 72 GTPLEPHEAKOYLYSEETLVANGSRTEOVGIYLLSSGD--GAQHRDSSGSKSRKRR 129  
 DB 29 GVPKKAEMNPQTSVNTGKREDAATKNOT-----SKAQVSNAPYIGTOKTSS----- 75  
 QY 130 IYGYDSRF-----SIFGKD-----FLNYPSTSVKST-----GCTGL 164  
 DB 76 LYGGCTELEKNIOTLQPSISITGDERRIRISSTFSFPRATVQSLIKYPTNSSTYGCGFL 135  
 QY 165 VAEKHVLTNAHCIIH-----DGKTYVGTOKLRVGLKPKFKQGGAND 208  
 DB 136 VNPNTVTVAGHCYVSDQHGMASTITAAPGRGSSYPYGIY----- 175  
 QY 209 STSNAPKPKQWIKRTHVPRKGMIGN---ANDIGMDYDALLKPKHKKRKNKIGV 265  
 DB 176 -SGTFEYVSK-GWTESKPTNDYGAIKNGSPGTVGW-YGRTTNSSP-----VGL 225  
 QY 266 SPPAKOLPGGRITFSGYDNDPRGNLYVRFCDYKDETVDLLYQCDAPGASGSGVYRMW 325  
 DB 226 SSSVIGFPCDKRTFGTMSDTRPIR-----SAETKLTLY-TDVFYGCQSGSPVY----- 272

OY 326 KRQOKWKRKIIIGFSGHWDANGSPQDENVAVRITPLKAQICW 372  
Db 273 -RMSDITGOTAIHRT-----NGC-SSYNGTRVTDVFNNIQW 310

RESULT 3  
ID ELL\_BOVIN STANDARD; PRT; 266 AA.

AC Q28153;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE ELASTASE 1 PRECURSOR (EC 3.4.21.36).

GN ELA1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
NCBI\_TaxID=9913;

RP SEQUENCE FROM N.A.  
RC STRAIN-HOUSTEIN-FRIESIAN; TISSUE-Pancreas;  
MEDLINE=98079203; PubMed=9418008;

RA Gestin M., le Huereu-Luron I., Wicker-Planquart C., le Drean G.,  
Chail J.C., Puigserver A., Guilloteau P.;  
"Bovine pancreatic preproelastases I and II: comparison of nucleotide  
RT and amino acid sequences and tissue specific expression.";  
RL Comp. Biochem. Physiol. 118B:181-187(1997).

CC -1- FUNCTION: ACTS UPON ELASTIN.  
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS, INCLUDING ELASTIN.  
CC -1- SUBCELLULAR LOCATION: SECRETED.

CC -1- TISSUE SPECIFICITY: PANCREAS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRIPSTIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: M80038; AAA98525.1; -.  
CC HSSP: P00772; TEST.

DR MEROPS: S01.153; -.  
DR InterPro: IPR001254; -.  
DR InterPro: IPR001314; -.  
DR Pfam: PF00089; trypsin.1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PROSITE: PS00134; TRYPSIN\_HIS.1.  
DR PROSITE: PS00135; TRYPSIN\_SER.1.

DR KW Hydrolyase; Serine protease; Zymogen; Pancreas; Signal.  
FT SIGNAL 1  
FT PROPEP 17 26  
FT CHAIN 27 266  
FT DISULFID 56 72  
FT DISULFID 153 220  
FT DISULFID 184 200  
FT DISULFID 210 240  
FT ACT\_SITE 71 71  
FT ACT\_SITE 119 119  
FT ACT\_SITE 214 214  
FT CARBOHYD 87 87  
FT CARBOHYD 241 241  
SQ SEQUENCE 266 AA; 28518 MW; 927E29C69BAF67E7 CRC64;

Query Match 6.0%; Score 124; DB 1; Length 266;  
Best Local Similarity 27.3%; Prol. No. 0.0026;  
Matches 35; Conservative 25; Mismatches 38; Indels 30; Gaps 6;

OY 147 NYPESTSVKLTG-----CGTLVAEKHVLTAACHIDGFTYVKGQKLKRVGLKPREK 200  
Db 37 SWPQSIISLYKSGSSWHTCGTILIKQKWMYTAACHVDSQWTF-----RVILG----- 84

OY 201 DGGKANDSTSAPEQKMFQIRKRTYVPGKWKGNANDIGMDVALLEL-KKPKRK 259  
Db 85 DHNLSQNDGTE-----QYISVQKIVVHPSW---NSNNVAGYDIALVRLAQSATLNS 133

OY 260 FMKIGVSP 267  
Db 134 YVQLGVLP 141

RESULT 4  
ID FA10\_RABIT STANDARD; PRT; 490 AA.

AC 019045;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE COAGULATION FACTOR X PRECURSOR (EC 3.4.21.6) (STUART FACTOR).  
GN F10.

OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
NCBI\_TaxID=9986;

RP SEQUENCE FROM N.A.  
RX MEDLINE=97256311; PubMed=9101642;  
RA Pundurthi U.R., Anderson K.D., James H.L.;  
RT "Characterization of a full-length cDNA for rabbit factor X.";  
RL Thromb. Res. 85:503-514(1997).

CC -1- FUNCTION: FACTOR XA IS A VITAMIN K-DEPENDENT GLYCOPROTEIN THAT  
CC CONVERTS PROTHROMBIN TO THROMBIN IN THE PRESENCE OF FACTOR VA,  
CC CA<sup>++</sup>, AND PHOSPHOLIPID DURING BLOOD CLOTTING.

CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-|-THR AND THEN  
CC ARG-|-ILE BONDS IN PROTHROMBIN TO FORM THROMBIN.

CC -1- SUBUNIT: THE TWO CHAINS ARE FORMED FROM A SINGLE-CHAIN PRECURSOR  
CC BY THE EXCISION OF TWO ARG RESIDUES AND ARE HELD TOGETHER BY 1 OR  
CC MORE DISULFIDE BONDS.

CC -1- PTM: THE VITAMIN K-DEPENDENT, ENZYMAIC CARBOXYLATION OF SOME  
CC GLUTAMIC ACID RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND  
CC CALCIUM (BY SIMILARITY).

CC -1- PTM: N- AND O-GLYCOSYLATED (BY SIMILARITY).  
CC -1- PTM: THE ACTIVATION PEPTIDE IS CLEAVED BY FACTOR IXA (IN THE  
CC INTRINSIC PATHWAY), OR BY FACTOR VIIA (IN THE EXTRINSIC PATHWAY)  
CC (BY SIMILARITY).

CC -1- MISCELLANEOUS: CALCIUM ALSO BINDS, WITH STRONGER AFFINITY TO  
CC ANOTHER SITE, BEYOND THE GLA DOMAIN.

CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRIPSTIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AF003200; AAB62542.1; -.  
CC InterPro: IPR000152; -.  
CC InterPro: IPR0001294; -.  
CC InterPro: IPR0000561; -.  
CC InterPro: IPR001254; -.  
CC InterPro: IPR001314; -.  
CC InterPro: IPR001881; -.  
CC InterPro: IPR002383; -.  
CC Pfam: PF00008; EGF.2.  
CC Pfam: PF00594; gla.1.  
CC Pfam: PF00089; trypsin.1.

Query Match 6.0%; Score 124; DB 1; Length 266;  
Best Local Similarity 27.3%; Prol. No. 0.0026;  
Matches 35; Conservative 25; Mismatches 38; Indels 30; Gaps 6;

```

DR PRINTS; PRO0001; GLABLOOD.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR PROSITE; PS00010; ASX_HYDROXYL. 1.
DR PROSITE; PS00011; GLU_CARBONYLATION. 1.
DR PROSITE; PS00022; EGF-1; 1.
DR PROSITE; PS01186; EGF-2; 2.
DR PROSITE; PS01187; EGF-CA; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Glycoprotein; Hydrolyase; Serine protease; Plasma; blood coagulation;
Gamma-carboxyglutamic acid; Hydroxylation; Calcium-binding; Vitamin K;
Signal; Zymogen; EGF-like domain; Repeat.
FT SIGNAL; 1 20
FT PROPEP 21 40
FT CHAIN 41 180
FT CHAIN 184 490
FT PROPEP 184 232
FT CHAIN 233 490
FT DOMAIN 86 122
FT DOMAIN 125 165
FT DOMAIN 233 490
FT MOD_RES 46 46
FT MOD_RES 47 47
FT MOD_RES 54 54
FT MOD_RES 56 56
FT MOD_RES 59 59
FT MOD_RES 60 60
FT MOD_RES 65 65
FT MOD_RES 66 66
FT MOD_RES 69 69
FT MOD_RES 72 72
FT MOD_RES 75 75
FT MOD_RES 79 79
FT MOD_RES 103 103
FT ACT_SITE 274 274
FT ACT_SITE 320 320
FT ACT_SITE 417 417
FT DISULFID 90 101
FT DISULFID 95 110
FT DISULFID 112 121
FT DISULFID 129 140
FT DISULFID 136 149
FT DISULFID 151 164
FT DISULFID 172 340
FT DISULFID 239 244
FT DISULFID 259 275
FT DISULFID 388 402
FT DISULFID 413 441
FT CARBOHYD 61 61
FT CARBOHYD 187 187
FT CARBOHYD 205 205
FT SEQUENCE 490 AA; 53965 MW; 3A39FA85AF2A6D11 CMC64;

Query Match 5.8%; Score 121.5; DB 1; Length 490;
Best Local Similarity 33.3%; Pred. No. 0.0088;
Matches 32; Conservative 15; Mismatches 24; Indels 25; Gaps 4;
OY 160 CTGTLVAEKHVLTAAHCCHDGTYYKGTQKLVGFLAKRFKDGAGANDSISAMEQMKF 219
DB 259 CGGTLSEVHVLTAAHCCHDGTYYKGTQKLVGFLAKRFKDGAGANDSISAMEQMKF 309

```

```

OY 220 QMTRVAKRTVHPKGIKGNANDIGMDYALLEKKP 255
DB 310 NRR-VKEIY-----DFDIAVLRLKTP 329

RESULT 5
ID EL2_PIG STANDARD; PRT; 269 AA.
AC P08419;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ELASTASE 2 PRECURSOR (EC 3.4.21.71).
GN ELA2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87217962; PubMed=3646943;
RA Kawashima I., Tani T., Shimoda K., Takiguchi Y.;
RA Shimada Y., Ozawa T., Miyayama T., Iijima H., Ishida A., Sato Y.,
RA Tama I., Tanaka J., Ikenaga H.;
RT "Molecular cloning and expression in Escherichia coli of a cDNA
RT encoding human pancreatic elastase 2."
RL J. Biochem. 102:1555-1563(1987).
CC [2]
CC MEDLINE=88198076; PubMed=2834346;
CC Shitazu Y., Yoshida H., Matsuki S., Takemura K., Ikeda N.,
CC Shimada Y., Ozawa T., Miyayama T., Iijima H., Ishida A., Sato Y.,
CC Tama I., Tanaka J., Ikenaga H.;
CC "Characterization of pancreatic elastase II cDNAs: two elastase II
CC mRNAs are expressed in human pancreas."
CC J. Biochem. 102:1555-1563(1987).
CC [1]
CC FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: LEU-I-XAA, MET-I-XAA
CC AND PHE-I-XAA. HYDROLYSES ELASTIN.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL; M16651; AAA31027.1; -
CC EMBL; D00237; BAA00166.1; -
CC PIR; A26823; A26823.
CC DR MEROPS; S01.135; -
CC DR InterPro; IPR001254; -
CC DR InterPro; IPR001314; -
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PRO0722; CHYMOTRYPSIN.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Pancreas; Zymogen; Signal.
FT SIGNAL 1 16
FT PROPEP 17 28
FT CHAIN 29 269
FT DISULFID 58 74
FT DISULFID 155 222
FT DISULFID 186 202
FT DISULFID 212 243
FT ACT_SITE 73 73
FT ACT_SITE 121 121
FT ACT_SITE 216 216
FT CONFLICT 10 10
L -> S (IN REF. 2).

```

FT CONFLICT 118 118 N -> K (IN REF. 2).  
FT CONFLICT 132 132 S -> Y (IN REF. 2).  
FT CONFLICT 172 172 I -> V (IN REF. 2).  
FT CONFLICT 202 202 C -> V (IN REF. 2).  
SQ SEQUENCE 269 AA: 28699 MW: BAC6FE69AF4DDE56 CAC64:

Query Match 5.7%; Score 119; DB 1; Length 269;  
Best Local Similarity 27.5%; Pred. NO. 0.0069;  
Matches 60; Conservative 28; Mismatches 64; Indels 66; Gaps 14;

QY 128 ROICGYDSRFISFGKDFLLNPFSTSVKL-STG-----CTGTLVAEKHYLAHCIDHDK 181  
DB 28 RVVGGEDARPN-----SMPWQVSLQYDSSGQMRHTCGGLTVQSVLTAHCISSSR 79  
QY 182 TY--VKGQKRLRVGFLKRFKFDGSGANDSTSAMPEQKFKQIRKRVKRVKGVKIGKMAN 239  
DB 80 TYRVVLGRHSL-----STNEPGLA-----VKVSKLVHODW---NSN 114  
DB 240 DIGMDYDVALLELKKP-HKRRFMKIGVSPPAKQ-LPG-----GRHFSGYDND--R 286  
DB 115 QLSNGNDIALKLKASPVSLTQKICLPAAGTILPNNYVCVYGTGWRHQTNGASPDILQ 174  
QY 287 PGNLYRRCVDEYDLYQCCDAPGASGSGVYVRM 324  
DB 175 OGOLL-----VVD-----YATC-SKPGMGSTVKTNM 200

## RESULT 6

EL2A\_HUMAN STANDARD; PRT; 269 AA.  
AC P08217;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 08, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ELASTASE 2A PRECURSOR (EC 3.4.21.71).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87217962; PubMed=3646943;  
RA Kawashima I., Tan T., Shimoda K., Takiguchi Y.;  
RT "Characterization of pancreatic elastase II cDNAs: two elastase II  
RT mRNAs are expressed in human pancreas.";  
RT DNA 6:163-172(1987).  
RL [2]

RP SEQUENCE FROM N.A.  
RX MEDLINE=88107669; PubMed=3427074;  
RA Fletcher T.S., Shen W.F., Lajman C.;  
RT "Primary structure of human pancreatic elastase 2 determined by  
RT sequence analysis of the cloned mRNA.";  
RL Biochemistry 26:7256-7261(1987).  
CC -1- FUNCTION: ACTS UPON ELASTIN.  
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: LEU-I-XAA, MET-I-XAA  
CC AND PHE-I-XAA. HYDROLYSES ELASTIN.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- TISSUE SPECIFICITY: PANCREAS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPsin FAMILY. ELASTASE SUBFAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: M16631; AAA52374.1; -;  
DR EMBL: M16652; AAA52380.1; -;  
DR PIR: A27432; A27432.

DR PIR: B26823; B26823.  
DR MEROPS: S01.155; -;  
DR InterPro: IPR001254; -;  
DR InterPro: IPR001314; -;  
DR Pfam: PF00089; trypsin.1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PROSITE: PS00134; TRYPsin\_HIS.1.  
DR PROSITE: PS00135; TRYPsin\_SER.1.  
KW Hydrolyase; Serine protease; Pancreas; zymogen; signal.

FT SIGNAL 1 16  
FT PROPEP 17 28  
FT CHAIN 29 269  
FT DISULFID 58 74  
FT DISULFID 155 222  
FT DISULFID 186 202  
FT DISULFID 212 243  
FT ACT\_SITE 73 73  
FT ACT\_SITE 121 121  
FT ACT\_SITE 216 216  
SQ SEQUENCE 269 AA: 28888 MW: A2E05143EF4987C CAC64;  
ACTIVATION PEPTIDE.  
ELASTASE 2A.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
BY SIMILARITY.  
CHARGE RELAY SYSTEM (BY SIMILARITY).  
CHARGE RELAY SYSTEM (BY SIMILARITY).

Query Match 5.6%; Score 115.5; DB 1; Length 269;  
Best Local Similarity 28.6%; Pred. NO. 0.014;  
Matches 50; Conservative 21; Mismatches 51; Indels 53; Gaps 11;

QY 147 NYPFSTSVKLTSG-----CTGTLVAEKHYLAHCIDHDKYVVGTOKLRVGFKPKFX 200  
DB 39 SMPWQVSLQYDSSGQMRHTCGGLTNSWVLTAAHCISSSRY-----RVGL----- 85  
QY 201 DGGRC-----ANDSTSAMPEQKFKQIRKRVKRVKGVKIGKMANDIGMDYDVALLELKKP- 255  
DB 86 -GRHNLVVAESGLA-----VSVSKIYVHKDW---NSNQISKNDIALKLAPV 131  
QY 256 HKRKMFKIGVSPPAKQ-LPG-----GRHFSGYDND--RPNLV---YRC 295  
DB 132 STFDKIQIACLPAGTILPNNYVCVYGTGWRHQTNGASPDILQ 186

## RESULT 7

EL2\_MOUSE STANDARD; PRT; 271 AA.  
AC P05208;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE ELASTASE 2 PRECURSOR (EC 3.4.21.71).  
GN ELA2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87066713; PubMed=3641189;  
RA Stevenson B.J., Hagenbuechle O., Wellauer P.K.;  
RT "Sequence organisation and transcriptional regulation of the mouse  
RT elastase II and trypsin genes.";  
RL Nucleic Acids Res. 14:8307-8330(1986).  
CC -1- FUNCTION: ACTS UPON ELASTIN.  
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: LEU-I-XAA, MET-I-XAA  
CC AND PHE-I-XAA. HYDROLYSES ELASTIN.  
CC -1- SUBCELLULAR LOCATION: SECRETED.  
CC -1- TISSUE SPECIFICITY: PANCREAS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPsin FAMILY. ELASTASE SUBFAMILY.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC -----

RL Blood 94:621-631 (1999).  
-i- FUNCTION: FACTOR XA IS A VITAMIN K-DEPENDENT GLYCOPROTEIN THAT  
CC -i- FUNCTION: FACTOR XA IS A VITAMIN K-DEPENDENT GLYCOPROTEIN THAT

CC CONVERTS PROTHROMBIN TO THROMBIN IN THE PRESENCE OF FACTOR VA  
CC CALCIUM, AND PHOSPHOLIPID DURING BLOOD CLOTTING. ACTS AS A TOXIN  
CC IN VENOM,  
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-1-THR AND THEN  
CC ARG-1-LIE BONDS IN PROTHROMBIN TO FORM THROMBIN.  
CC -1- SUBUNIT: THE TWO CHAINS ARE FORMED FROM A SINGLE-CHAIN PRECURSOR  
CC AND ARE HELD TOGETHER BY 1 OR MORE DISULFIDE BONDS.  
CC -1- PM: THE VITAMIN K-DEPENDENT, ENZYMAIC CARBOXYLATION OF SOME  
CC GLUTAMIC ACID RESIDUES ALLOWS THE MODIFIED PROTEIN TO BIND  
CC CALCIUM.  
CC -1- PM: THE ACTIVATION PEPTIDE IS CLEAVED BY FACTOR IXA (IN THE  
CC INTRINSIC PATHWAY), OR BY FACTOR VIIA (IN THE EXTRINSIC PATHWAY)  
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY.  
CC MEROPS; S01.216; -  
DR InterPro: IPR000152; -  
DR InterPro: IPR000294; -  
DR InterPro: IPR000561; -  
DR InterPro: IPR001254; -  
DR InterPro: IPR001314; -  
DR InterPro: IPR001436; -  
DR InterPro: IPR001881; -  
DR InterPro: IPR002383; -

DR PRINTS; PRO0010; LSC4000;  
DR PRINTS; PRO0722; CHYMOTRYPSIN.  
DR PRINTS; PS00010; ASX\_HYDROXYL; 1.  
DR PROSITE; PS00011; GLU\_CARBOXYLATION; 1.  
DR PROSITE; PS00011; GLU\_CARBOXYLATION; 1.

```
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01188; EGF_CA; 1.
```

```
DR PROSITE; PS00134; TRYPSPIN_SER; 1.  
DR PROSITE; PS00135; TRYPSPIN_SER; 1.
```

KW Glycylproline, hyaluronate, organic phosphorus  
KW Gamma-carboxylglutamic acid; Hydroxylation; Calcium-binding; Vitamin  
KW Gamma-irradiation; non-ionic domain; Repeat

	FACTOR X LIGHT CHAIN.
CHAIN	141
NON CONS	142
FT	141
FT	141

FT CHAIN	142	376	FACTOR X HEAVY CHAIN.
ET CHAIN	142	376	ACTIVATED FACTOR XA, HEAVY CHAIN.

FT	DOMAIN	50	81	EGF-LIKE 1, CALCIUM-BINDING.
FT	DOMAIN	89	124	EGF-LIKE 2.

FT	DOMAIN	142	376	CATALYTIC.
FT	MOD_RES	6	6	GAMMA-CARBOXYGLUTAMIC ACID.

FTT	/	GAMMA-CARBOXYGLUTAMIC ACID.
MOD_RES	14	GAMMA-CARBOXYGLUTAMIC ACID.
MOD_RES	14	GAMMA-CARBOXYGLUTAMIC ACID.
FTT	/	GAMMA-CARBOXYGLUTAMIC ACID.

[illegible]

F1	MOD_RES	26	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	25	GAMMA-CARBOXYGLUTAMIC ACID.
MOD_PTC	MOD_PTC	26	GAMMA-CARBOXYGLUTAMIC ACID.

FT	MOD_RES	29	GAMMA-CARBOXYGLUTAMIC ACID.
EM	MOD_PEC	32	GAMMA-CARBOXYGLUTAMIC ACID.

FT	35	GAMMA-CARBOXYGLUTAMIC ACID.
MOD_RES	63	HYDROXYLATION (BY SIMILARITY).
MOD_RES	35	
MOD_RES	63	

FT	ACT_SITE	183	CHARGE RELAY SYSTEM (BI SIMILARITY).
FT <td>ACT_SITE <td>229 <td>CHARGE RELAY SYSTEM (BY SIMILARITY). </td></td></td>	ACT_SITE <td>229 <td>CHARGE RELAY SYSTEM (BY SIMILARITY). </td></td>	229 <td>CHARGE RELAY SYSTEM (BY SIMILARITY). </td>	CHARGE RELAY SYSTEM (BY SIMILARITY).

FT	ACT_SITE	326	326	CHARGE RELAY SYSTEM (BI SIMILARITY).
FT	DISULFID	17	22	BY SIMILARITY.

FT	50	61	BI SIMILANITY.
DISULFID	55	70	BY SIMILARITY.
FT			BI SIMILANITY.

FT	DISULFID	12	01	DI	SIMILARITY
FT	DISULFID	89	100	BY	SIMILARITY.
FT	DISULFID	89	100	BY	SIMILARITY.

FT	123	INTERCHAIN (BY SIMILARITY).
DISULFID	111	BY SIMILARITY.
DISOLF ID	124	
F1	105	

BY SIMILARITY.	
DISULFID	148
DICURED	168
BY SIMILARITY.	184
DISULFID	153
FT	148
F1 DISULFID	153

DISINFID	297	BY SIMILARITY.
DISINFID	322	BY SIMILARITY.
DISINFID	350	BY SIMILARITY.

FT	CARBOHYD	52	52
----	----------	----	----



FT CARBOHYD 186 186 N-LINKED (GLCNAc...)  
SQ SEQUENCE 376 AA; 42455 MM; F5A5C7FE9BA561C1 CRC64;

Query Match 5.5%; Score 115; DB 1; Length 376;  
Best Local Similarity 23.8%; Pred. No. 0.022; Mismatches 72; Indels 72; Gaps 11;  
Matches 54; Conservative 29;

QY 66 GPOC-----HKG-----TPLLTYE--EAKOYLSTETLYANGS-----RTE 98  
DB 47 GDGSSNCHRGTCCKDGTGTCCTCLPNECKNEKVLXGSCARDNGNCHMFCRQVSE 106  
QY 99 TQVGI---YIISSSGDGAQHNRDSSGSKS--RRKQIYGYDSRFSIFGKDFLNTPFST 153  
DB 107 TQSCAESYRGLGVGDHSCVAEGDFSCGRNIRKARKNIVNGMDCK-----LGECPPQAV 158  
QY 154 VKLSTG---CTGTVAEKHYLTAACHIDCKTYVKGCTKLRGFLKPKFGKGRANST 210  
DB 159 LINEKGEVFCGGTILSPIHLVLAHCINQTKS-VKETRRL----- 197  
QY 211 SAMPEOMFOWIRVKTVPKRG--WIKGNANDIGMDYDALLEKRP 255  
DB 198 -----LSVDKRYVHTKFPVPPYVYVHQNPDRAVDIATIRMTTP 238

## RESULT 9

TRY5\_AEDAE STANDARD; PRT: 238 AA.

ID TRY5\_AEDAE STANDARD; PRT: 238 AA.  
AC P29787;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE TRYPsin 561 PRECURSOR (EC 3.4.21.4) (FRAGMENT).  
OS Aedes aegypti (yellowfever mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
OC Culicidae; Aedes.  
OX NCBI\_Taxid=7159;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97242546; PubMed=9087545;  
RA Kalkok S., Tabak L.M., Prosser D.E., Brooker W., Downer A.E.R.,  
RA White B.N.;  
RT "Isolation, sequencing and characterization of two cDNA clones coding  
for trypsin-like enzymes from the midgut of Aedes aegypti.";  
RL Insect Mol. Biol. 2:71-79(1993).  
-1- FUNCTION: MAJOR FUNCTION MAY BE TO AID IN DIGESTION OF THE BLOOD  
MEAL.  
-1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
-1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
-1- TISSUE SPECIFICITY: MIDGUT.  
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
TRYPsin FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL: X64363; CAA45715.1;  
DR PIR: S19891; TRWV5Y.  
DR HSSP: P00763; IDPO.  
DR MEROPS: S01.112;  
DR InterPro: IPR001254;  
DR Pfam: PF00089; Trypsin. 1.  
DR PROSITE: PS00134; TRYPsin\_HIS. 1.  
DR PROSITE: PS00135; TRYPsin\_SER. 1.  
RM Hydrolyase; Serine protease; zymogen; digestion; Multigene family.  
FT NON\_TER 1  
FT PROPEP <1 11 ACTIVATION PEPTIDE.

FT CHAIN 12 238 TRYPsin 561.  
FT ACT\_SITE 53 53 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 97 97 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT DISULFID 38 54 BY SIMILARITY.  
FT DISULFID 162 178 BY SIMILARITY.  
FT DISULFID 189 213 BY SIMILARITY.  
FT SITE 187 187 REQUIRED FOR SPECIFICITY (BY SIMILARITY).  
SQ SEQUENCE 238 AA; 25269 MM; 3BA22FE2EA32E4B5 CRC64;

Query Match 5.5%; Score 114.5; DB 1; Length 238;  
Best Local Similarity 29.4%; Pred. No. 0.014;  
Matches 42; Conservative 22; Mismatches 44; Indels 35; Gaps 7;

QY 149 PFTS---VKLTGCTGLVAEKHYLTAACHIDCKTYVKGCTKLRGFLKPKFGKGRG 205  
DB 24 PFOVSLGCVGSHFGGSLSERWMTAGHCAASQTKL---QVRIG--SSQHASGG-- 75  
QY 206 ANDSTSAMPEOMFOWIRVKTVPKRG--WIKGNANDIGMDYDALLEKRPKRFRKIG 264  
DB 76 -----QLIKVKVNRHPR-----YDEVTTDYDFALLEETVTPSDSCAP 115  
QY 265 VSPPAKQLP---GGRHFGCTDN 284  
DB 116 VKLPQKDTFVNEGTCLQVSGWGN 138

## RESULT 10

EIL\_PIG STANDARD; PRT: 266 AA.

ID EIL\_PIG STANDARD; PRT: 266 AA.  
AC P00772; Q29625;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-APR-1986 (Rel. 07, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE ELASTASE 1 PRECURSOR (EC 3.4.21.36).  
GN ELA1.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
OX NCBI\_Taxid=9823;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86304235; PubMed=3528137;  
RA Shitrasu Y., Yoshida H., Miyayama T., Matsuki S., Tanaka J.,  
RA Ikenaga H.;  
RT "Isolation and expression in Escherichia coli of a cDNA clone  
encoding porcine pancreatic elastase.";  
RL J. Biochem. 99:1707-1712(1986).  
RN (2)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87250343; PubMed=3648024;  
RA Tani T., Kawashima I., Furukawa H., Ohmine T., Takiguchi Y.;  
RT "Characterization of a silent gene for human pancreatic elastase I:  
structure of the 5'-flanking region.";  
RL J. Biochem. 101:591-599(1987).  
RN (3)  
RP SEQUENCE OF 27-266.  
RX MEDLINE=73229121; PubMed=4578945;  
RA Shotton D.M., Hartley B.S.;  
RT "Evidence for the amino acid sequence of porcine pancreatic  
elastase.";  
RL Biochem. J. 131:643-675(1973).  
RN (4)  
RP SEQUENCE OF 27-266.  
RX MEDLINE=70114042; PubMed=5415108;  
RA Shotton D.M., Hartley B.S.;  
RT "Amino-acid sequence of porcine pancreatic elastase and its  
homologies with other serine proteinases.";  
RL Nature 225:802-806(1970).  
RN (5)  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=78111463; PubMed=628010;

SAWYER L., SHOTTON D.M., CAMPBELL J.W., WENDELL P.L., MUIRHEAD H.,  
 WATSON H.C., DIAMOND R., LADNER R.C.;  
 "The atomic structure of crystalline porcine pancreatic elastase at  
 2.5-A resolution: comparisons with the structure of  
 alpha-chymotrypsin."  
 J. Mol. Biol. 118:137-208(1978).  
 [6]  
 X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS), AND ACTIVE SITE.  
 MEDLINE:70114044; PUBMED:5415110;  
 SHOTTON D.M., WATSON H.C.;  
 "Three-dimensional structure of tosyl-elastase."  
 Nature 225:811-816(1970).  
 [7]  
 X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF COMPLEX WITH INHIBITOR.  
 MEDLINE:95006335; PUBMED:7922044;  
 HANG K., STRYNAKA N.C., BERNARD V.D., PEANASKY R.J., JAMES M.N.;  
 "The molecular structure of the complex of Ascaris  
 chymotrypsin/elastase inhibitor with porcine elastase."  
 Structure 2:679-689(1994).  
 CC -1- FUNCTION: ACTS UPON ELASTIN.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS, INCLUDING ELASTIN.  
 CC PREFERENTIAL CLEAVAGE: ALA-I-XAA.  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- TISSUE SPECIFICITY: PANCREAS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPSIN FAMILY.  
 CC -1- DATABASE: NME-worthington enzyme manual;  
 WWW="http://www.worthington-biochem.com/manual/E/ES.html".  
 -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----  
 DR EMBL: X04036; CAZ2670.1;  
 DR EMBL: D00070; BAA00049.1;  
 DR EMBL: D00160; BAA00118.1;  
 DR PIR: JSD013; ELPG.  
 DR PDB: 1EST; 16-OCT-87.  
 DR PDB: 2EST; 30-APR-94.  
 DR PDB: 3EST; 15-OCT-92.  
 DR PDB: 4EST; 15-JUL-92.  
 DR PDB: 5EST; 15-APR-92.  
 DR PDB: 6EST; 15-OCT-91.  
 DR PDB: 7EST; 30-APR-94.  
 DR PDB: 8EST; 15-JAN-93.  
 DR PDB: 9EST; 31-JAN-94.  
 DR PDB: 11MC; 31-JAN-94.  
 DR PDB: 11JC; 31-JAN-94.  
 DR PDB: 1ELA; 30-APR-94.  
 DR PDB: 1ELB; 22-JUN-94.  
 DR PDB: 1ELC; 30-APR-94.  
 DR PDB: 1ELD; 14-FEB-95.  
 DR PDB: 1ELE; 14-FEB-95.  
 DR PDB: 1ELF; 10-JUL-95.  
 DR PDB: 1ELG; 10-JUL-95.  
 DR PDB: 1ESA; 30-APR-94.  
 DR PDB: 1ESB; 30-APR-94.  
 DR PDB: 1EAT; 05-APR-99.  
 DR PDB: 1EAS; 07-FEB-95.  
 DR PDB: 1EAT; 07-FEB-95.  
 DR PDB: 1EAV; 07-FEB-95.  
 DR PDB: 1EAW; 07-DEC-95.  
 DR PDB: 1LIV; 27-JAN-97.  
 DR PDB: 1NES; 29-JAN-96.  
 DR PDB: 1BOE; 18-NOV-98.  
 DR PDB: 1BTU; 16-FEB-99.  
 DR MEROPS: S01.153;  
 DR InterPro: IPR001254;  
 DR InterPro: IPR001314;

DR Pfam: PF00089; trypsin. 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PROSITE: PS00134; TRYPSIN\_HIS. 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER. 1.  
 KM Hydrolyase; Serine protease; 3D-structure; zymogen; pancreas; signal.  
 FT SIGNAL 1 16  
 FT PROPE 17 26  
 FT CHAIN 27 72  
 FT DISULFID 56 72  
 FT DISULFID 153 220  
 FT DISULFID 184 200  
 FT DISULFID 210 240  
 FT ACT\_SITE 71 71  
 FT ACT\_SITE 119 119  
 FT ACT\_SITE 214 214  
 FT CONFLICT 92 92  
 FT CONFLICT 126 126  
 FT CONFLICT 184 184  
 FT CONFLICT 204 204  
 FT STRAND 28 28  
 FT TURN 29 29  
 FT STRAND 31 32  
 FT TURN 35 36  
 FT TURN 39 40  
 FT STRAND 41 48  
 FT TURN 49 50  
 FT STRAND 51 62  
 FT TURN 63 64  
 FT STRAND 65 68  
 FT HELIX 70 72  
 FT STRAND 79 83  
 FT STRAND 87 87  
 FT STRAND 96 105  
 FT TURN 107 108  
 FT TURN 111 112  
 FT TURN 114 116  
 FT HELIX 114 125  
 FT STRAND 121 125  
 FT TURN 143 144  
 FT TURN 149 150  
 FT STRAND 153 157  
 FT STRAND 160 160  
 FT TURN 163 164  
 FT STRAND 167 167  
 FT STRAND 170 170  
 FT STRAND 172 175  
 FT STRAND 178 179  
 FT HELIX 181 184  
 FT TURN 185 185  
 FT TURN 187 190  
 FT TURN 191 193  
 FT HELIX 191 197  
 FT TURN 196 197  
 FT STRAND 198 201  
 FT STRAND 208 208  
 FT TURN 211 212  
 FT TURN 214 215  
 FT STRAND 217 222  
 FT STRAND 223 224  
 FT TURN 225 234  
 FT STRAND 225 234  
 FT TURN 236 236  
 FT TURN 237 238  
 FT STRAND 239 239  
 FT TURN 243 244  
 FT STRAND 247 251  
 FT STRAND 252 254  
 FT HELIX 252 255  
 FT TURN 255 255  
 FT HELIX 256 263  
 FT TURN 264 265  
 SO SEQUENCE 266 AA; 28821 MW; BF07D6855BB50FE2 CRC64;

Query Match 5.4%; Score 113; DB 1; Length 266;  
 Best Local Similarity 27.3%; Pred. No. 0.022;  
 Matches 36; Conservative 21; Mismatches 41; Indels 34; Gaps 6;

```

OY 147 NPSTSVKLTSG-----CTGTLYAEKHYLTAACIHDKTY--VKGTOKLRVGLAKP 198
DB 37 SWPQSISSLOYSRSGSWAHCTGTLIRQWVMTAAHCVDRELFRRVVGHHN-----88
OY 199 FPDGGRGANDSTSAPEQMKFQWIRKRVKPKWIKGANDIGMDYDALLEL-KKPK 257
DB 89 -----NNDGTE-----OYGVOKIVHPYV---NTDDVAGYDIALRLAOSVTL 131
OY 258 RKFMRIGVSPPA 269
DB 132 NSYVOLGVLPRA 143

RESULT 11
ELI_RAT STANDARD: PRT: 266 AA.
AC P00773:
21-JUL-1986 (Rel. 01, Created)
21-JUL-1986 (Rel. 01, Last sequence update)
15-JUL-1999 (Rel. 38, Last annotation update)
DE ELASTASE 1 PRECURSOR (EC 3.4.21.36).
GN ELA1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82182967; PubMed=6918221;
RA McDonald R.J., Swift G.H., Quinto C., Swain W., Pictet R.L.,
RA Nikovits W., Rutter W.D.;
RT "Primary structure of two distinct rat pancreatic preproelastases
RT determined by sequence analysis of the complete cloned messenger
RT ribonucleic acid sequences."
RL Biochemistry 21:1453-1463(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85054882; PubMed=6094548;
RA Swift G.H., Craik C.S., Stary S.J., Quinto C., Lahaie R.C.,
RA Rutter W.J., McDonald R.J.;
RT "Structure of the two related elastase genes expressed in the rat
RT pancreas."
RL J. Biol. Chem. 259:14271-14278(1984).
RN [3]
RP SEQUENCE OF 17-45.
RT TISSUE-Pancreas;
RL MEDLINE=84000385; PubMed=6555050;
RA Lajman C.;
RT "Isolation and characterization of rat pancreatic elastase."
RL Biochemistry 22:3763-3770(1983).
CC -1- FUNCTION: ACTS UPON ELASTIN.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS, INCLUDING ELASTIN.
CC PREFERENTIAL CLEAVAGE: ALA-|-XAA.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: PANCREAS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. ELASTASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01234; CA24544.1; -
DR EMBL: L00117; AAA98811.1; -
DR EMBL: L00112; AAA98811.1; JOINED.
DR EMBL: L00113; AAA98811.1; JOINED.
DR EMBL: L00114; AAA98811.1; JOINED.
DR EMBL: L00115; AAA98811.1; JOINED.

```

```

DR EMBL: L00116; AAA98811.1; JOINED.
DR PIR: A00960; ELRT1.
DR HSSP: P00772; 1EST.
DR MEROPS: S01.153; -.
DR InterPro: IPR001254; -.
DR InterPro: IPR001314; -.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; trypsin; 1.
DR PROSITE: PS00134; TRYPsin_HIS; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Pancreas; zymogen; signal.
FT SIGNAL 1 16
FT PROPEP 17 26
FT CHAIN 27 266
FT DISULFID 56 72
FT DISULFID 153 220
FT DISULFID 184 200
FT DISULFID 210 240
FT ACT_SITE 71 71
FT ACT_SITE 119 119
FT ACT_SITE 214 214
FT CONFLICT 104 104
FT CONFLICT 108 108
FT CONFLICT 244 244
FT CONFLICT 266 266
SQ SEQUENCE 266 AA; 28976 MW; 5A56FE8FCF1AAEDA CRC64;

```

Query Match 5.4%; Score 113; DB 1; Length 266;  
 Best Local Similarity 26.7%; Pred. NO. 0.022; Mismatches 48; Indels 36; Gaps 7;  
 Matches 39; Conservative 23;

```

OY 128 ROIVGYDSRFSIFGKDFLLNYPSTSVKLTSG-----CTGTLYAEKHYLTAACIHDKT 182
DB 26 RYVGAEARBNWSPQISLQY-----LSGSGWYHTGCGTLIRQWVMTAAHCVSSQMT 78
OY 183 YVKGTOKLRVGLAKPKFKDGGANDSTSAPEQMKFQWIRKRVKPKWIKGANDIG 242
DB 79 F-----RVVVG-----DHNLSQDGT-----OYVSOKIVHPYV---NSNVA 115
OY 243 MDYDALLEL-KKPKRKFMRIGVSP 267
DB 116 AGYDIALRLAOSVTLNNYVOLAVLP 141

RESULT 12
TRYP_STRGR STANDARD: PRT: 259 AA.
ID TRYP_STRGR
AC P00775;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TRYPSIN PRECURSOR (EC 3.4.21.4) (SGT).
GN SPRT.
OC Streptomyces griseus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1911;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 10137;
RL MEDLINE=92095977; PubMed=1755852;
RA Kim J.C., Cha S.H., Jeong S.T., Oh S.K., Byun S.M.;
RT "Molecular cloning and nucleotide sequence of Streptomyces griseus
RT trypsin gene."
RL Biochem. Biophys. Res. Commun. 181:707-713(1991).
RN [2]
RP SEQUENCE OF 37-259.
RX MEDLINE=75127940; PubMed=804314;
RA Olafson R.W., Jurasek L., Carpenter M.R., Smillie L.B.;
RT "Amino acid sequence of Streptomyces griseus trypsin. Cyanogen
RT bromide fragments and complete sequence."
RL Biochemistry 14:1168-1177(1975).

```



DR	PIR:	A25853;	A25853.
DR	PIR:	A24478;	A24478.
DR	PDB:	1HCG; 08-MAY-95.	
DR	PDB:	1FAX; 29-OCT-97.	
DR	PDB:	1FXV; 17-JUN-98.	
DR	PDB:	1KKA; 23-MAR-99.	
DR	PDB:	1XKB; 23-MAR-99.	
DR	MEROPS:	S01.216; -.	
DR	GlycoSuiteDB:	P00742; -.	
DR	CarbBank:	CCSD:29393; -.	
DR	MIM:	134530; -.	
DR	MIM:	227600; -.	
DR	InterPro:	IIPR00152; -.	
DR	InterPro:	IIPR00294; -.	
DR	InterPro:	IIPR00561; -.	
DR	InterPro:	IIPR01254; -.	
DR	InterPro:	IIPR01314; -.	
DR	InterPro:	IIPR01881; -.	
DR	InterPro:	IIPR02383; -.	
DR	Pfam:	PF00098; EGF_2.	
DR	Pfam:	PF00594; gla; 1.	
DR	Pfam:	PF00089; trypsin; 1.	
DR	PRINTS:	PRO0001; GLABLOD.	
DR	PRINTS:	PRO0722; CHYMOTRYPSIN.	
DR	PROSITE:	PS00010; ASX_HYDROXYL; 1.	
DR	PROSITE:	PS00011; GLU_CARBOXYLATION; 1.	
DR	PROSITE:	PS00022; EGF_1; 1.	
DR	PROSITE:	PS00134; TRYPSIN_HIS; 1.	
DR	PROSITE:	PS00135; TRYPsin_SER; 1.	
DR	PROSITE:	PS01186; EGF_2; 2.	
DR	PROSITE:	PS01187; EGF_CA; 1.	
KW	Glycoprotein:	Hydrolase; Serine protease; Plasma; Blood coagulation;	
KW	Gamma-carboxyglutamic acid:	Hydroxylation; Calcium-binding; Vitamin K	
KM	Signal:	Zymogen; EGF-like domain; Repeat; 3D-structure.	
FT	SIGNAL	1	?
FT	PROPEP	? 40	
FT	CHAIN	41 179	FACTOR X LIGHT CHAIN.
FT	CHAIN	183 488	FACTOR X HEAVY CHAIN.
FT	PROPEP	183 234	ACTIVATION PEPTIDE.
FT	CHAIN	235 488	ACTIVATED FACTOR XA, HEAVY CHAIN.
FT	DOMAIN	86 122	EGF-Like 1, CALCIUM-BINDING (POTENTIAL)
FT	DOMAIN	125 163	EGF-Like 2.
FT	DOMAIN	235 488	CATALYTIC.
FT	MOD_RES	46 46	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	47 47	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	54 54	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	56 56	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	59 59	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	60 60	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	65 65	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	66 66	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	69 69	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	72 72	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	79 79	GAMMA-CARBOXYGLUTAMIC ACID.
FT	MOD_RES	103 103	GAMMA-CARBOXYGLUTAMIC ACID.
FT	CARBOHYD	199 199	N-LINKED (GLCNAC. . .).
FT	CARBOHYD	211 211	N-LINKED (GLCNAC. . .).
FT	CARBOHYD	221 221	CHARGE RELAY SYSTEM.
FT	CARBOHYD	231 231	CHARGE RELAY SYSTEM.
FT	ACT_SITE	276 276	CHARGE RELAY SYSTEM.
FT	ACT_SITE	322 322	CHARGE RELAY SYSTEM.
FT	ACT_SITE	419 419	CHARGE RELAY SYSTEM.
FT	DISULFID	90 101	
FT	DISULFID	95 110	
FT	DISULFID	112 121	
FT	DISULFID	129 140	
FT	DISULFID	136 149	
FT	DISULFID	151 164	
FT	DISULFID	172 342	
FT	DISULFID	241 246	
FT	DISULFID	261 277	
FT	DISULFID	390 404	
FT	DISULFID	415 443	
INTERCHAIN.			

FT CONFLICT 285 288 KVRV -> E (IN REF. 4 AND 6).  
 FT CONFLICT 442 442 G -> S (IN REF. 3).  
 SQ SEQUENCE 488 AA; 54731 MM; F81D5746AF4797AF CRC64;

Query Match  
 Best Local Similarity 24.4%; Score 111.5; DB 1; Length 488;  
 Matches 49; Conservative 28; Mismatches 59; Indels 63; Gaps 10;

DB 160 CTGTLVAEKHYLAACIHGKTYVKTOKLRYGFLKPKKDGGRANDSTAMPQMKF 219  
 261 CGGTLLSEFLLTAHCLYAKRF-----KVRGDRNTGEGEGEAVHE-VEVYIKHNR 314  
 220 QWIKRTHYPKWIKGNANDIGMDYALLETLEKPKHKKFMKIGVSPPA----- 269  
 315 -----TKETY-----DFDIALVRLKTPIT---FRNNVAPACLPERDMAEST 352  
 270 --KOLPG-----GRHESGYDNDPRPNLYRFDCKVD-----ETYDLLYQ 307  
 353 LMTQKTGIVSGFGRTHKGRSTRKMLKLEVPYVDNRNSCKLSSEFLLTQNNMFCAGYDT--K 410  
 308 QCDAPGASGSGVYRMWK 326  
 411 QEDACQGDGSG-GRHYTRK 428

RESULT 14  
 NETR\_HUMAN STANDARD; PRT; 875 AA.

AC 356730; 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE NEUROTRYPsin PRECURSOR (EC 3.4.21.-) (MOTOPsin).  
 GN PRESS12.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98201705; PubMed=9540828;  
 RA Proda K., Geschwend P.P., Sonderegger P.;  
 RT "Cloning and sequencing of the cDNA encoding human neurolysin.";  
 RL Biochim. Biophys. Acta 1396:143-147(1998).  
 CC -1- FUNCTION: PLAYS A ROLE IN NEURONAL PLASTICITY AND THE PROTOLYTIC  
 CC ACTION MAY SUBSERVE STRUCTURAL REORGANIZATIONS ASSOCIATED WITH  
 CC LEARNING AND MEMORY OPERATIONS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: SECRETED.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE  
 CC TRYPsin FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 KRINGLE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 4 SRCR DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AJ001531; GAA04816.1; -  
 DR InterPro: IPR000001; -  
 DR InterPro: IPR001190; -  
 DR InterPro: IPR001254; -  
 DR InterPro: IPR001314; -  
 DR PRINTS: PR00258; SPERACTRCPTR.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR PROSITE: PS00420; SPERACT\_RECEPTOR; 3.

KW Hydrolyase; Serine protease; Glycoprotein; Kringle; Repeat; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 875  
 FT DOMAIN 23 92  
 FT DOMAIN 93 165  
 FT DOMAIN 170 271  
 FT DOMAIN 280 381  
 FT DOMAIN 387 487  
 FT DOMAIN 500 601  
 FT DOMAIN 619 875  
 FT DOMAIN 630 631  
 FT ACT\_SITE 630 631  
 FT ACT\_SITE 676 676  
 FT ACT\_SITE 726 726  
 FT ACT\_SITE 825 825  
 FT DISULFID 619 750  
 FT CARBOHYD 26 26  
 FT CARBOHYD 683 683  
 SQ SEQUENCE 875 AA; 97011 MM; B66EC946DC208DC8 CRC64;

Query Match  
 Best Local Similarity 23.1%; Score 111.5; DB 1; Length 875;  
 Matches 73; Conservative 47; Mismatches 85; Indels 111; Gaps 21;

DB 118 SGSSGKS-----RRKQIYGYDSRFSIGKDFLL-NYPFSYVKLSTG----- 159  
 607 SGNSKESLSVSGGLRLRRQKRIIG-----GKNSLRGGMPOVSLRLKSSHGGR 658  
 160 --CTGTLVAEKHYLAACIHGKTYVKTOK--LRVGLRKKFKFDGGRANDSTAMPE 215  
 659 ILGCAVLLSSCWLLTAHCF---KRYGNSTRSYAVRAG-----DYITLYPE 701  
 216 QMK-----FQWIRYKTRHYPKWIKGNANDIGMDYALLETLEKPKHKKFMKIGVSPPA 269  
 702 EFEELGVQOIVYHREYRD-----RSDYDIALVRLGPEHCARFSS-HYLPAC 750  
 270 -----KOLPGRIHPSGY-DNDPRGN-----LYRFDCKVD---TYDLL--- 305  
 751 LPLMRERPKTASNCYITGWDGTRAYSRLLQOALPLPKRCERKRYKGRFTGRMLCAG 810  
 306 -----YQCCDAOPGASGSGVYRMWKROQKWEKIIIGFSHOWVDMNGSPQDFNVAARI 361  
 811 NLEHKKVNDSCQDSSGSL---MCEPREGSW--VYGVIS---W-----GYGGVAD 854  
 362 TPLKYAQICY---WIK 374  
 855 SPGYTVSAFVPMIK 870

RESULT 15  
 CFAD\_HUMAN STANDARD; PRT; 253 AA.

AC P00746;  
 DT 21-JUN-1986 (Rel. 01, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE COMPLEMENT FACTOR D PRECURSOR (EC 3.4.21.46) (C3 CONVERTASE ACTIVATOR)  
 DE (PROPERDIN FACTOR D) (ADIPSIN).  
 GN DF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE OF 8-253 FROM N.A.  
 RX MEDLINE=92250520; PubMed=1374388;  
 RA White R.T., Damm D., Hancock N., Rosen B.S., Lowell B.B., Usher P.,  
 RA Flier J.S., Spiegelman B.M.;  
 RT "Human adipisin is identical to complement factor D and is expressed  
 RT at high levels in adipose tissue";  
 RL J. Biol. Chem. 267:9210-9213(1992).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RA Filler J.S., Spiegelman B.M., Rosen B.M.;  
 RL Patent number WO9006365, 14-JUN-1990.  
 RN [3]  
 RP SEQUENCE OF 26-252.  
 RX MEDLINE-85000441; PubMed-6383466;  
 RA Nleman M.A., Bhown A.S., Bennett J.C., Volanakis J.E.;  
 RT "Amino acid sequence of human D of the alternate complement  
 pathway."  
 RL Biochemistry 23:2482-2486(1984).  
 RN [4]  
 RP PARTIAL SEQUENCE OF 26-252.  
 RX MEDLINE-84108950; PubMed-6363133;  
 RA Johnson D.M.A., Gagnon J., Reid K.B.M.;  
 RT "Amino acid sequence of human factor D of the complement system.  
 Similarity in sequence between factor D and proteases of non-plasma  
 origin."  
 RL FEBS Lett. 166:347-351(1984).  
 RN [5]  
 RP PARTIAL SEQUENCE OF 26-61 AND 194-220.  
 RX MEDLINE-84256515; PubMed-6821372;  
 RA Johnson D.M.A., Gagnon J., Reid K.B.M.;  
 RT "Factor D of the alternative pathway of human complement.  
 Purification, alignment and N-terminal amino acid sequences of the  
 major cyanogen bromide fragments, and localization of the serine  
 residue at the active site."  
 RL Biochem. J. 187:863-874(1980).  
 RN [6]  
 RP PARTIAL SEQUENCE OF 26-82.  
 RX MEDLINE-80145719; PubMed-6987665;  
 RA Volanakis J.E., Bhown A.S., Bennett J.C., Mole J.E.;  
 RT "Partial amino acid sequence of human factor D: homology with serine  
 proteases."  
 RL Proc. Natl. Acad. Sci. U.S.A. 77:1116-1119(1980).  
 RN [7]  
 RP PARTIAL SEQUENCE OF 26-78.  
 RX MEDLINE-81054886; PubMed-6776531;  
 RA Davis A.E. III;  
 RT "Active site amino acid sequence of human factor D."  
 RL Proc. Natl. Acad. Sci. U.S.A. 77:4938-4942(1980).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE-94118317; PubMed-8289289;  
 RA Narayana S.V.L., Carson M., El-Kabbani O., Kilpatrick J.M., Moore D.,  
 Chen X., Bugb C.E., Volanakis J.E., Delucas L.J.;  
 RT "Structure of human factor D. A complement system protein at 2.0-A  
 resolution."  
 RL J. Mol. Biol. 235:695-708(1994).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE-96025834; PubMed-7592653;  
 RA Kim S., Narayana S.V., Volanakis J.E.;  
 RT "Crystal structure of a complement factor D mutant expressing  
 enhanced catalytic activity."  
 RL J. Biol. Chem. 270:24399-24405(1995).  
 CC -1- FUNCTION: FACTOR D CLEAVES FACTOR B WHEN THE LATTER IS COMPLEXED  
 WITH FACTOR C3B, ACTIVATING THE C3BB COMPLEX, WHICH THEN BECOMES  
 THE C3 CONVERTASE OF THE ALTERNATE PATHWAY. ITS FUNCTION IS  
 HOMOLOGOUS TO THAT OF C1S IN THE CLASSICAL PATHWAY.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 TRYPsin FAMILY.  
 CC -1- CAUTION: IN ADDITION TO THE CONFLICTS SHOWN IN THE FEATURE TABLE,  
 REF.3 SEQUENCE HAD A PEPTIDE IN THE WRONG ORDER AND ANOTHER ONE  
 MISSING.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL; M84526; AAA35527.1; ALT\_INIT.  
 DR PIR; A40197; DBHU.  
 DR PDB; 1DEF; 25-FEB-98.  
 DR PDB; 1DST; 11-JUL-96.  
 DR PDB; 1DSU; 17-AUG-96.  
 DR MEROPS; S01.191; -.  
 DR MIM; 134350; -.  
 DR InterPro; IPR001254; -.  
 DR InterPro; IPR001314; -.  
 DR Pfam; PF00089; trypsin; 1.  
 DR PRINTS; PR00722; CHYMOTRYPSIN.  
 DR PROSITE; PS00134; TRYPsin\_HIS; 1.  
 DR PROSITE; PS00135; TRYPsin\_SER; 1.  
 KW Complement alternate pathway; Plasma; Hydrolase; Serine protease;  
 KZ Zymogen; Glycoprotein; Signal; 3D-structure.  
 FT SIGNAL 1 20  
 FT PROPEP 1 25  
 FT CHAIN 26 253  
 FT ACT\_SITE 66 66  
 FT ACT\_SITE 112 112  
 FT ACT\_SITE 208 208  
 FT DISULFID 51 67  
 FT DISULFID 148 214  
 FT DISULFID 179 195  
 FT DISULFID 204 229  
 FT CONFLICT 26 26  
 FT CONFLICT 35 35  
 FT CONFLICT 40 40  
 FT CONFLICT 49 49  
 FT CONFLICT 52 52  
 FT CONFLICT 59 59  
 FT CONFLICT 63 63  
 FT CONFLICT 73 73  
 FT CONFLICT 83 86  
 FT CONFLICT 83 84  
 FT CONFLICT 94 95  
 FT CONFLICT 96 96  
 FT CONFLICT 136 136  
 FT CONFLICT 178 191  
 FT CONFLICT 243 243  
 FT CONFLICT 250 250  
 FT CONFLICT 250 250  
 SQ SEQUENCE 253 AA; 27004 MW; BD53B70BD5C6AD CRC64;  
 Query Match 5.3%; Score 111; DB 1; Length 253;  
 Best Local Similarity 28.4%; Pred. No. 0.03; Mismatches 43; Indels 48; Gaps 9;  
 Matches 44; Conservative 20;  
 QY 149 PFSTSVKLTSG--CTGLVAEKHVLTAACIH--DGKTYVKGTOKLKRVGFLKPKFKDGG 203  
 DB 38 PYMASVQNLGAHLCGVLAEGQWVLSAAHCEADADGQVYL----- 79  
 QY 204 RGANDSTSNPEQMKFQWIRKRVKTHVPRKGIKGNANDIGMDYALLLEKKPKRKPKKI 263  
 DB 80 LGAHLSLSPPEPSKRYLDVLAFA---VP---HPESQPTIDHLLDLQLSE-----KA 124  
 QY 264 GVSPPAKQLPGGRIFHSQYDND-RPGNLYVRCDV 297  
 DB 125 TLGPAVRPLPMQRY-----DRDVAPGTL-----CDV 150

Search completed: July 12, 2001, 11:44:23  
 Job time: 107 sec

Thu Jul 12 15:25:01 2001

us-09-554-933-3.rsp

Page 14



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 12, 2001, 11:41:55 ; Search time 24.68 Seconds  
(without alignments)  
2053.194 Million cell updates/sec

Title: US-09-554-933-3  
Perfect score: 2080  
Sequence: 1 MAGIPGLFLFLCAVQ.....LKYAQICWIKGNLDCREG 383

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues  
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SPREMBL\_16:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_proteus:\*  
12: sp\_unclassified:\*  
13: sp\_vertebrate:\*  
14: sp\_virus:\*

.Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2080	100.0	383	4	095084
2	129.5	6.2	303	2	09EXR9
3	129	6.2	469	6	09GMD9
4	127.5	6.1	1322	5	09NATO
5	126.5	6.1	678	11	09JUS8
6	124.5	6.0	1322	5	09NUS5
7	124	6.0	266	6	046644
8	123.5	5.9	339	11	09OX91
9	123.5	5.9	366	11	09OX85
10	123.5	5.9	541	11	09OX90
11	123.5	5.9	623	11	09JUP3
12	123.5	5.9	643	11	09OX84
13	122	5.9	259	5	09X161
14	122	5.9	449	5	09VDO8
15	120.5	5.8	482	11	063207
16	118	5.7	1376	5	09VOR8
17	117	5.6	258	5	09W508
18	117	5.6	522	5	001771
19	115.5	5.6	269	4	014243

20	115	5.5	258	4	09UN11	09un11 homo sapien
21	113.5	5.5	418	5	09VAB7	09vab7 drosophila
22	113.5	5.5	685	11	092338	092338 mus musculu
23	113.5	5.5	737	13	090422	090422 brachydanto
24	113	5.4	573	5	09V516	09v516 drosophila
25	112.5	5.4	273	2	P76176	P76176 escherichia
26	112.5	5.4	603	3	09VDV1	09vdv1 drosophila
27	112	5.4	249	13	09W7Q1	09w7q1 paralichthy
28	110.5	5.3	279	5	096991	096991 manduca sex
29	110.5	5.3	481	11	054740	054740 mus musculu
30	110	5.3	1449	5	010922	010922 caenorhabdi
31	109.5	5.3	405	2	069973	069973 streptomyce
32	108	5.2	274	5	017086	017086 anopheles s
33	108	5.2	750	13	09W633	09w633 cyprinus ca
34	108	5.2	1047	5	024019	024019 drosophila
35	107.5	5.2	236	11	09Z1H1	09z1h1 mus musculu
36	107.5	5.2	481	11	088947	088947 mus musculu
37	107.5	5.2	611	5	09VZ58	09vz58 drosophila
38	107	5.1	603	11	061129	061129 mus musculu
39	106	5.1	460	5	09VFP6	09vfp6 drosophila
40	106	5.1	762	13	09VIC6	09vic6 cyprinus ca
41	105.5	5.1	268	13	09W7Q2	09w7q2 paralichthy
42	105.5	5.1	604	11	09WUW3	09wuw3 rattus norv
43	105	5.0	787	5	09VER6	09ver6 drosophila
44	104.5	5.0	290	5	09VRT2	09vrt2 drosophila
45	104	5.0	1190	5	09V7B5	09v7b5 drosophila

## ALIGNMENTS

RESULT 1  
ID 095084 PRELIMINARY; PRT; 383 AA.  
AC 095084:  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
DE SERINE PROTEASE (HYPOTHETICAL 43.0 KDA PROTEIN).  
GN ZS1G13 OR DKFP586B0719.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_Taxid:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=UMBILICAL VEIN;  
RA Li X., Tedder T.F.,  
RT "A novel serine protease from human umbilical vein endothelial  
RT cells.";  
RL submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Sheppard P., Blumberg H., Jelinek L., Foster D., O'Hara P.,  
RL submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP TISSUE=UTERUS;  
RC SEQUENCE FROM N.A.  
RA Mambutt R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.,  
RL submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
DR EMBL: AF015287; ADO1553.1; -;  
DR EMBL: AF193611; AAF07186.1; -;  
DR EMBL: AL136914; CAB66848.1; -;  
DR MEROPS: S01.309; -;  
DR InterPro: IPR001254; -;  
DR InterPro: IPR001314; -;  
DR Pfam: PF00089; trypsin.1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR SMART: SM00020; TRYP\_SPC; 1.  
KW Hydrolase; Protease; Serine protease.

SEQUENCE 383 AA; 43001 MW; 46EB6C11ABFDE58F CRC64;

Query Match 100.0%; Score 2080; DB 4; Length 383;  
Best Local Similarity 100.0%; Pred. No. 2e-181;  
Matches 383; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAGIPGLFLFLFLCAVGVSPYSAFWKPTWAVRLPVLPSTLNLAKPDEGAAKLE 60  
DB 1 MAGIPGLFLFLFLCAVGVSPYSAFWKPTWAVRLPVLPSTLNLAKPDEGAAKLE 60  
QY 61 VSSSGGQCHKGTEPLPYEEAKOYLSYETLYANGSRTETGVGITYLLSSGDAQHDSGS 120  
DB 61 VSSSGGQCHKGTEPLPYEEAKOYLSYETLYANGSRTETGVGITYLLSSGDAQHDSGS 120  
QY 121 SGKSRRRKQIYGYDSRFSIFGKDFLNPSTSTYKLTGCTGLVAKHVLTAHCHDQ 180  
DB 121 SGKSRRRKQIYGYDSRFSIFGKDFLNPSTSTYKLTGCTGLVAKHVLTAHCHDQ 180  
QY 181 KTYVKGTOKLVGFLEKPKFKGSGANDSTSAMPEQKFWIRVYKRTVHFKWIKGNAND 240  
DB 181 KTYVKGTOKLVGFLEKPKFKGSGANDSTSAMPEQKFWIRVYKRTVHFKWIKGNAND 240  
QY 241 IGMDDYVALLLEKPKFKKFKKFKKFKKFKKFKKFKKFKKFKKFKKFKKFKKFKK 300  
DB 241 IGMDDYVALLLEKPKFKKFKKFKKFKKFKKFKKFKKFKKFKKFKKFKKFKKFKK 300  
QY 301 TYDLYQOCDAOPGASGGVYVYRWKROQKWKERKIIIGIFSGHGVDMNGSPQDENAVR 360  
DB 301 TYDLYQOCDAOPGASGGVYVYRWKROQKWKERKIIIGIFSGHGVDMNGSPQDENAVR 360  
QY 361 ITPLYKAOICWIKGNVLDRCRG 383  
DB 361 ITPLYKAOICWIKGNVLDRCRG 383

RESULT 2  
Q9EXR9 PRELIMINARY; PRT: 303 AA.

AC Q9EXR9; 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, last sequence update)  
DE 01-MAR-2001 (TREMblrel. 16, last annotation update)  
DE GLUTAMYL-ENDOPEPTIDASE.  
OS Bacillus intermedius.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_Taxid=1400;  
RN [1]  
RP SEQUENCE FROM N.A.  
MEDLINE=99170141; PubMed=10071925;  
REBrikov D.V., Akimkina T.V., Shevchey A.B., Demiduk I.V.,  
Bushman A., Kostrov S.V., Chestukhina G.G., Stepanov V.M.,  
"Bacillus intermedius glutamyl endopeptidase. Molecular cloning and  
RT nucleotide sequence of the structural gene.";  
J. Protein Chem. 18:21-25(1999).  
DR EMBL: Y15136; CAC17594.1;  
SQ SEQUENCE 303 AA; 32343 MW; F2E37B5B07781D59 CRC64;

Query Match 6.2%; Score 129.5; DB 2; Length 303;  
Best Local Similarity 24.6%; Pred. No. 0.00088;  
Matches 65; Conservative 40; Mismatches 98; Indels 61; Gaps 15;

QY 108 SSGGAGHRDSSGSGKRRKQIYGYDSRFSIFGKDFLNPSTSTYKLTG--CTGL 164  
DB 68 SSGGAGHRDSSGSGKRRKQIYGYDSRFSIFGKDFLNPSTSTYKLTG--CTGL 164  
QY 165 VAEKHLVLAACIHD--GKTY-VKGTOKLVGFLKPKDGKSGANDSTSAMPEQKFW 221  
DB 125 IAPKILILNGHCYVNTASRSYSAKGS-----YYP-----GMDSTAVNGS----- 164  
QY 222 IIVKRTVHPKWTIGNANDIGMDYVALLLEKPKFKKFKKFKKFKKFKKFKKFKK 275

DB 165 ANMTEFYVPSGYINTGAS-----QYDAVITD-----TNINQTYGYSIROVYMLTGT 213  
QY 276 RIHFGSYDND--RPGNLVYRF--CDVKEDEYDLYQOCDAOPGASGGVYVYRWKROQ 330  
DB 214 TIKISGYGDAMRSTGKVKSGWEMSGSVTRBEDTNLAYTYIDTFSGNSGSA---MLDONOQ 269  
QY 331 KWERTIIGIF-SGHQWDMNGSPQ 353  
DB 270 ----IVGVHNGYSNGTINGPK 288

RESULT 3

Q9GMD9 PRELIMINARY; PRT: 469 AA.

AC Q9GMD9; 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, last sequence update)  
DE 01-MAR-2001 (TREMblrel. 16, last annotation update)  
DE COAGULATION FACTOR X.  
OS Ornithorhynchus anatinus (Duckbill platypus).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Monotremata; Ornithorhynchidae; Ornithorhynchus.  
OX NCBI\_Taxid=9258;  
RN [1]  
RP SEQUENCE FROM N.A.  
Poorafshar M.M., Hellman L.L.;  
RT Identification and structural analysis of three serine proteases in a  
monotreme, platypus, Ornithorhynchus anatinus.";  
Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
DR EMBL: AF275654; AAG00453.1;  
SQ SEQUENCE 469 AA; 52196 MW; 4C66C230D0758F6A CRC64;

Query Match 6.2%; Score 129; DB 6; Length 469;  
Best Local Similarity 26.5%; Pred. No. 0.0018;  
Matches 66; Conservative 30; Mismatches 73; Indels 80; Gaps 16;

QY 160 CTGLVAKHVLTAHCHDGTYYKGTOKLVGFLKPKFKGSGANDSTSAMPEQKMF 219  
DB 258 CGGTILNEYIILSAHCHMOAKRF-----KVRGGERDTEKKSSEMAHEKVIYHS-KF 311  
QY 220 QMIRVYKTHPKWIKGNANDIGMDYVALLLEKPKFKKFKKFKKFKKFKKFKKFKK 270  
DB 312 ----VKRTY-----DEDIIVIKIKTPTIT--FRMNVSPACLEKMDAEDI 349  
QY 271 ----QLPG-----GRHFGSYDNDPGLN-----YRFDVKDETYDLYQO-C----- 309  
DB 350 LMNQKAGVYSGFGVHEKRG-----RPSVILKMLEVYPERVETTCROSSSFDTIPMFCAGTD 405  
QY 310 ----DAOPGASGGVYVYRWKROQKWKERKIIIGIFSGHGVDMNGSPQDENAVRITPLK 365  
DB 406 SRPDAQOGDSG-GPHVTKYKDY-----FYIGIVMGEGCAQNGK--FGVYT-----K 451  
QY 366 YAOICWYWK 374  
DB 452 AATFLSNWK 460  
RESULT 4  
Q9NATO PRELIMINARY; PRT: 1322 AA.  
AC Q9NATO; 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, last sequence update)  
DE 01-MAR-2001 (TREMblrel. 16, last annotation update)  
DE ADHESIVE SERINE PROTEASE.  
GN SP22D.  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
OC Anophelinae.  
OX NCBI\_Taxid=7165;

[1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE-20318993; PubMed-10860981;  
 RA Dantelli A., Loukeris T., Lagneux M., Mueller H.M., Richman A.,  
 RA Kafatos F.C.;  
 RT "A modular chitin-binding protease associated with hemocytes and  
 RT hemolymph in the mosquito *Anopheles gambiae*.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:7136-7141(2000).  
 CC -1 SIMILARITY: TO LOW DENSITY LIPOPROTEIN (LDL) RECEPTOR CLASS A  
 CC (LDLRA) DOMAIN.  
 CC -1 SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
 DR EMBL: AJ276428; CAB81934.1; -.  
 DR InterPro: IPR001190; -.  
 DR InterPro: IPR001254; -.  
 DR InterPro: IPR001314; -.  
 DR InterPro: IPR002172; -.  
 DR InterPro: IPR002557; -.  
 DR Pfam: PF00057; 1dl\_recept\_a; 2.  
 DR Pfam: PF00089; trypsin; 1.  
 DR Pfam: PF00089; trypsin; 1.  
 DR Pfam: PF00530; SRCR; 2.  
 DR Pfam: PF01607; Chitin\_bind\_2; 2.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PRINTS: PR00261; LDLRECEPTR.  
 DR PRINTS: PR00258; SPERACTRPT.  
 DR PROSITE: PS01209; LDLRA\_1; 1.  
 DR PROSITE: PS00068; LDLRA\_2; 2.  
 DR PROSITE: PS00420; SRCR\_1; UNKNOWN\_1.  
 DR PROSITE: PS0287; SRCR\_2; 2.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR SMART: SM00494; Chtd2; 1.  
 DR GlycoProtein: Protease.  
 KW SEQUENCE 1322 AA; 146794 MW; FBF973C21CC5475B CRC64;

Query Match 6.1%; Score 127.5; DB 5; Length 1322;  
 Best Local Similarity 25.3%; Pred. No. 0.0097;  
 Matches 46; Conservative 35; Mismatches 58; Indels 43; Gaps 9;

OY 128 ROIVYDGRFSIFGKDFLLNPESTSVKLT--GCTGLVNAKHYLTAAHCHDKTYVK 185  
 DB 1078 RYVHSE---TVYG-----HHWQASLRKTKMHWGAVLITRYHYLTAAHCLIG--YPK 1126  
 OY 186 GTOKLRVGLFKPKFKDGRGANDSTSAMPEQKFMQIRKTHVPGKVIKGNANDIGMDY 245  
 DB 1127 STYRVRIG-----DYHTAAYDNAELD-IFENYTHIEQREGH-----HMSN 1167  
 OY 246 DYALLELKKPKHKKRKKMGKIGVSPAPAK--LPGRIHFSGYDNDPRGNLVYRCDVKTET 302  
 DB 1168 DIAVAVVLTVPVRFNDYVOPICLPARDAPYLPQONCTISGWGTEAGS-----KDSY 1219  
 OY 303 DL 304  
 DB 1220 DL 1221  
 RESULT 5  
 O9JUS8 PRELIMINARY; PRT; 678 AA.  
 AC O9JUS8;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE MANNOSE-BINDING PROTEIN ASSOCIATED SERINE PROTEASE-2 PRECURSOR  
 DE (FRAGMENT).  
 GN MASP-2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wallis R., Dodd R.B.;

RT "Interaction of mannose-binding protein with associated serine  
 RT proteases: Effects of naturally occurring mutations.";  
 RT Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
 CC -1 SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
 DR EMBL: AJ277747; CAB90832.1; -.  
 DR InterPro: IPR000152; -.  
 DR InterPro: IPR000436; -.  
 DR InterPro: IPR000561; -.  
 DR InterPro: IPR000859; -.  
 DR InterPro: IPR001254; -.  
 DR InterPro: IPR001314; -.  
 DR InterPro: IPR001881; -.  
 DR Pfam: PF00084; sushl; 2.  
 DR Pfam: PF00089; trypsin; 1.  
 DR Pfam: PF00431; CUB; 2.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
 DR PROSITE: PS01180; CUB; 2.  
 DR PROSITE: PS01186; EGF\_2; 1.  
 DR PROSITE: PS01187; EGF\_CA; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR SMART: SM00032; CCP; 1.  
 KW Calcium-binding; EGF-like domain; Glycoprotein; Protease; Repeat;  
 KW Signal.  
 FT NON\_TER 1 1  
 FT SIGNAL <1 12 POTENTIAL.  
 FT CHAIN 13 678 MANNOSE-BINDING PROTEIN ASSOCIATED SERINE  
 FT PROTEASE-2.  
 SQ SEQUENCE 678 AA; 74734 MW; 16680E4A9ADCC05 CRC64;

Query Match 6.1%; Score 126.5; DB 11; Length 678;  
 Best Local Similarity 22.1%; Pred. No. 0.0049;  
 Matches 79; Conservative 46; Mismatches 123; Indels 109; Gaps 19;

OY 65 GCP-----QCHKPPLPYEBAKOYLSTLYANGSRTEOVGIYIISSGDGAQH 115  
 DB 359 GCPDPLPNGHVDYITGGEVTVYKAVIQYSCHEFTYTMSSN-----GRYVEADGFTWS 413  
 OY 116 RDS-----GSSGSRKRRQIYGYDSRSISGKFLNYPSTSVKLTGCTGLV 165  
 DB 414 GKEKSLPVCKPYGSLSTHTSGRIITGGAPKPG-----DEPMQVLLIGETAAAGALI 465  
 OY 166 AEKHYLTAAHCHDKTYVYKTKLRVGLFKPKFKDGRGANDSTSAMPEQKFMQIRK 225  
 DB 466 HDWVLTAAHAY-KTEAMSSLDTRMGITK-----RLSHYQANPEAV----- 509  
 OY 226 RTHVPGKVIKGNANDIGMDYDALLELRK--PHKRRKMGKIGVSPAPAKLPG-----GR 276  
 DB 510 --FIEHGYTHG---AGPDNDIALIKLKVKYININIMPICL--PRKEAASIMKTDVFGT 561  
 OY 277 IHFSGYDNDPRG---NLVYRCDVKTETDLYOQC-----DAQPPA----- 315  
 DB 562 V--AGWGLTKGKFLARNLMFVDIPVD-----HQKCATAYTKQDPYPAKYTVNNMLCAGL 613  
 OY 316 -----SSGSVYVRMKRROOKMERKIIIGFSGHGVDMN--GSPDQFNVAVRIT 362  
 DB 614 DAGKDSGRDSDGALVFLDNETQRM--FVGSIYS--WGSINGSGSRYGYTKVT 665  
 RESULT 6  
 O9NUS5 PRELIMINARY; PRT; 1322 AA.  
 AC O9NUS5;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE SERINE PROTEASE 22D.  
 GN SP22D.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Anopheles.

RA Kawashima I.,  
RL Submitted (Aug-1995) to the EMBL/GenBank/DDBJ databases.  
CC - - SIMILARITY TO SERINE PROTEASES, TRYPSIN FAMILY.  
CC - - SIMILARITY TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
DR EMBL; X91440; CAA62746.1; -.  
DR HSSP; P00772; 1EIG.  
DR InterPro; IPR001254; -.  
DR InterPro; IPR001314; -.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00722; CHYMOTRYPSIN.  
DR PROSITE; PS00134; TRYP SIN HIS; UNKNOWN\_1.  
DR PROSITE; PS00135; TRYP SIN\_SER; 1.  
DR SMART; SM00020; Tryp\_Spec; 1.Signal.  
KW Hydrolase; Serine protease; Signal.  
FT SIGNAL  
FT CHAIN 1 16 POTENTIAL.  
FT CHAIN 27 266 POTENTIAL.  
SO SEQUENCE 266 AA; 28754 MW; 6B3281B90F53A9IC CRC64;

	Query Match	6.0%;	Score 124;	DB 6;	Length 266;	
	Best Local Similarity	28.1%;	Pred.No. 0.0023;			
	Matches	36;	Conservative	24;	Mismatches 38;	Indels 30; Gaps 6;
QY	147 NYPSTSVKLTG-----CTGTVAEKHVLPAACIHDKGYVKGTOQLRVGHLPKRK	200	:::: :::	::  ::	:	:
Db	37 SMPQSISLQYLSGGSWHYHTGCGTLLIKONNVMAALACVSPKTF-----RYYVG-----	84				
QY	201 DQGRANDSTSAMDEQKPFOWIRFKRTTHPKGMKINGNDIGMDYDALLEL-KKPKRK	259				
Db	85 DHNISQNNGTE-----QYVSQKIIVHPYV--NSNNVAAGDIALLRLAOSTPLNS	133	::: :::	::  ::	:	:
QY	260 FMKIGVSP 267		::: :::			
Db	134 YVQLGVLP 141					

Query Match	Score	DB	Length
Best local Similarity	5.9%	123.5; DB 11;	339;
	22.1%;	Pred. No. 0.0036;	

Matches	79;	Conservative	46;	Mismatches	123;	Indels	109;	Gaps	19
QY	65	CGP-----	QCHKGPPLPYTEAKOYLSYETLYANGSRRTETOVGIYILSSSGDGAH	115					
	111								
Db	20	CGPPDPLNGHVDYITTCGEVTTTAKAVIQSCSETEFYTMSN-----	GKYVCEADGEFTSS	74					
QY	116	RDS-----	GSSGKSRRKKQIYGDSDRSITGKDFLWTFPSTSKLSTGCGTIV	165					
	116								
Db	75	KGEKSLPVCKPYCGSLSTHTSGRIITGGOPAKG-----	DFPMOYLLIGETTAAGALI	126					
QY	166	AEKHVLTAAHCCHIDKCTYVKGTFKRLVGFGLKPKFKDGGANDSTSAMPEOKKFWIRVK	225						
	127								
Db	127	HDDWVLTAAHAYV-GKTEAMSSLDLRMGILK-----	RLSLITYQAMPEAV-----	170					
QY	226	RTHVPRGMIKGNANDIGMDYDYLLELK--PHKRKFMTIGVSPPAQLPG-----	GR	276					
	171								
Db	171	-FIEHGTHG---AGFDNDIALIKLKNKVTINNRMPTCL--PRKEAASLMKTDVFCT	222						
QY	277	IHFSGYDNDREB-----NLVYRFCDYVKDETYLLYQOC-----	DAOPGA-----	315					
	223								
Db	223	V--AGWGLTGKGFARLNLMFDPIDPVD-----	HQCAVATYKOPYPGAKATVNMICAGL	274					
QY	316	-----SSGGVYVRMMKROOKMERKIIIGFSGHWDVNN--GSPDQFNFAVRIT	362						
	275	DRGKDCRCRGSBGALVFLDNETQRM--FVGIVS---WGSINCGSGSDQIYGYITKVT	326						

RESULT	9
ID	090x85
AC	PRELIMINARY; PRT; 366 AA.
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DR	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE	01-MAR-2001 (TREMBLrel. 16, Last annotation update)
GN	MAPI9. PROTEIN (FRAGMENT).
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxId=10116;
PN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=FISHER 344;
RA	Stover C.M., Thiel S., Lynch N.J., Schwaebel W.J.;
RT	"The Rat and mouse homologues of MASP-2 and MAP19, components of the
FT	Mannan-Binding Lectin Activation Pathway of Complement."
CD	Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.
CC	-1 STIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (51).
EMBL:	Y18571; CAB65388.1; -
DR	HSSP; P20231; IAAO.
DR	InterPro: IPR000436; -
DR	InterPro: IPR001254; -
DR	InterPro: IPR001314; -
DR	Pfam: PF00084; sushi; 2.
DR	Pfam: PF00089; trypsin; 1.
DR	PRINTS: PR00722; CHYMOTRYPSIN.
DR	PROSITE: PS00135; TRYPSIN_SER; 1.
DR	SMART: SM00032; CCP; 1.
FT	NON_TER
SO	SEQUENCE 366 AA; 39882 MW; F84F5ACF0EC93468 CRC64;

  

Query Match	5.9%; Score 123.5; DB 11; Length 366;
Best Local Similarity	22.1%; Pred. NO. 0.004;
Matches 79; Conservative	46; Mismatches 123; Indels 109; Gaps 19;

  

Qy	116	RDS-----GSGGRSRKROLYGVDSRSIFGKDFLNYPSTSVKLSTGCTGLV	165
Dz	102	KGEKSLPVCKPYCGISTHTSGRIITGGAPKPG-----DPEWQVILLGETTAAGALI	153

```

OY      166  A EKHVTAHCHIHQGTQKTVGTOKLRNGEFLPKFRKDGSRGNDSTSPMPOMKQWIRK  225
Db      154  H D D W V I T A A H A V Y - G K T A M S S L D I R M G I L K ----- R S L I Y T Q M P E A V -----  197
OY      226  R T H V P K G W I K N A N D I G M D Y D A L L E L K - - P H K R E M K I G V S P A K Q L P G - - - - - G R  276
Db      198  - F I H E G Y T H G - - - - A G F D N D I A L I K L K N K V I T N R I N M P I C L - - P R K E A S L M K T D P V G T  249
OY      277  I H F S Y D N D R G - - - - N L V Y F P C V K D E T Y D L L Y Q O C - - - - - D A Q G A - - - - -  315
Db      250  V - A G M G L T Q G F L A R N I M F D I P D I V D - - - - H O K A T A V T K Q P Y G A R V T N M L C A G L  301
OY      316  - - - - - S S G G V Y R M K R Q O O K W E R K I I G F S G H O W D M - G S P O D E N V A V R I T  362
Db      302  D R G K D C R G D S G G A L Y F L D E T O R W - F V G G I A S - - - - W S I N C G S E O Y G V I T K V T  353

```

RESULT	10
09QX90	
ID	09QX90 PRELIMINARY; PRT; 541 AA.
AC	09QX90;
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE	01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE	MASP-2 PROTEIN (FRAGMENT).
GN	MASP-2.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;	
[1]	
RP	SEQUENCE FROM N.A.
RC	STRAIN=FISHER 344;
RC	Stover C.M., Thiel S., Lynch N.J., Schwaebel W.J.;
RL	"The Rat and mouse homologues of MASP-2 and Map19, components of the
RL	Mannan-Binding lectin Activation Pathway of Complement.";
CC	Submitted (DEC-1998) to the EMBL/Genbank/DDBJ databases.
-I-	SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).
DR	EMBL; Y118566; CAB65383.1; -
DR	HSSP; P00736; IAPQ.
DR	InterPro: IPRO00152; -
DR	InterPro: IPRO000436; -
DR	InterPro: IPRO000561; -
DR	InterPro: IPRO000859; -
DR	InterPro: IPRO01254; -
DR	InterPro: IPRO01314; -
DR	Pfam; PF00084; sushi; 2.
DR	Pfam; PF00089; trypsin; 1.
DR	Pfam; PF00431; CUB; 1.
DR	PRINTS; PR00722; CHYMOTRYPSIN.
DR	PROSITE; PS0010; ASX_HYDROXYL; UNKNOWN_1.
DR	PROSITE; PS01180; CUB; 1.
DR	PROSITE; PS01186; ESGF_2; 1.
DR	PROSITE; PS00135; TRYPSIN_SER; 1.
DR	SMART; SM00032; CCP; 1.
KW	EGF-like domain; Glycoprotein.
FT	NON_TER
SEQUENCE	541 AA; 59323 MM; 51EBCC0C1A8549CA8 CRC64;

  

Query Match	5.9%; Score 123.5; DB 11; Length 541;
Best Local Similarity	22.1%; Pred. No. 0.0067;
Matches	79; Conservative 46; Mismatches 123; Indels 109; Gaps 19;

  

Qy	65 CGP-----GCKKGPPLPYEEAKROYLSLETLYANGSRRETOVGIIYLSSSGDAQH	115
Db	222 CGPPDLLPNGHAVDYITGVEVTVYKAVIDYSCEETFYTMSSN-----GRVCADDFWTSS	276
Qy	116 RDS-----GSSGSKRRKKOYIGVDSRSIFGDKDFLNVPYSTSVKLSTGCTGLV	165
Db	277 KGEKSLPVCKPYPCGSLSTHTSGRGRIIGGAPAKPG-----DFPOVULLGETTTAGALTI	328

RESULT	11		
09JUP3			
ID	09JUP3	PRELIMINARY;	PRY; 623 AA.
AC	09JUP3		
DT	01-OCT-2000	(TIREMBrel. 15, Created)	
ST	01-OCT-2000	(TIREMBrel. 15, last sequence update)	
MA	01-MAR-2001	(TIREMBrel. 16, last annotation update)	
DE	MA5P-2	PROTEIN (FRAGMENT)	

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus  
RX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER 344;  
RC MEDLINE=20054576; PubMed=10580686;  
RA Stover C.M., Thiel S., Lynch N.J., Schwaible W.J.;  
RT "The rat and mouse homologues of MASP-2 and Map19, components of the  
RT mannan-binding lectin activation pathway of complement.";  
RL J. Immunol. 163:6848-6859(1999).  
CC 1-1. SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
EBHL Y18564; CAB70973.1; "-  
DR InterPro: IPR000152; "-  
DR InterPro: IPR000436; "-  
DR InterPro: IPR000561; "-  
DR InterPro: IPR000589; "-  
DR InterPro: IPR001254; "-  
DR InterPro: IPR001314; "-  
DR InterPro: IPR001881; "-  
DR Pfam: PF00084; sushi; 2.  
DR Pfam: PF00089; trypsin; 1.  
DR Pfam: PF00431; CUB; 1.  
DR PRINTS: PRO0722; CHYMOTRYPSIN.  
DR PROSITE: PS00010; ASX\_HYDROXYL; UNKNOWN\_1.  
DR PROSITE: PS01180; CUB; 2.  
DR PROSITE: PS01186; EGF\_2; 1.  
DR PROSITE: PS01187; EGF\_CA; 1.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
DR SMART: SM00032; CCP; 1.  
DR Calcium-binding; EGF-like domain; Glycoprotein; Repeat.  
KM NON\_CATH\_1  
SQ SEQUENCE 623 AA; 68568 MW; 233AF39EB906624F CRC64;

**OY** 65 GCP-----QCHKSTPLTYTEAKQKSLVETLXANSRRETOVGILYLSSGGDAQH 115  
          |||               |||||  
**Db** 304 CSDPDLNPNHVDYITIGEVETIITNAVIQISCDETFITMSN-----GAYVEADGFWTSS 358  
                                |||  
**OY** 116 RDS-----GSCKSRKKRQIYGVSRFEIPECKDMLWYPSRSVKLSCTCGTGLV 165  
                                |||  
**Db** 359 KEKSLPYCKRPVCGISLTHSGRIITGCPARKG-----DFPMQVILLSETPAAGALI 410

RESULT	12	
Q9OX84	PRELIMINARY;	PTT: 643 AA.
ID	Q9OX84	
AC	Q9OX84	
DT	01-MAY-2000	(TREMBLrel, 13, Created)
DT	01-MAY-2000	(TREMBLrel, 13, Last sequence update)
DT	01-MAR-2001	(TREMBLrel, 16, Last annotation update)
DE	MAP19 PROTEIN (FRAGMENT).	

Query Match	5.9%	Score 123.5;	DB 11;	Length 643;
Best Local Similarity	22.1%;	Pred. No. 0.0085;		
Matches	79;	Conservative	46;	Mismatches 123;
				Indels 109;
				Gaps 19

QY 65 CGP-----QCHKGTPLPYEAKOYLSYETLYANGSRIFQVGYILSSGDAOH ILT  
||| | : || | || | | : : |

Dh 344 CGPDDLPNGHVDYITGEVTTYKAVIOYSCFEFFTYTSSN-----GRVCEADGEFTSS 378  
Oy 116 RDS-----GSSGSRKROIYGYDSRPSFGKDFLLNTPSTSVKLTGCTFLV 165  
Dh 379 KGENSLPCKPVCGLSTHTSSGRIGGOPARPG-----DFPQVILLGETTAGALI 430  
Oy 166 AEKHLVTAHCIDHGTQYVKTQKLVGFLPKPKDGRGANDSTSAPEQKQWIRVK 225  
Dh 431 HDDWVLTAAHAYV-GKTEAMSSLDIRMGILK-----RLSLITYQAMPEAV----- 474  
Oy 226 RTHVPKGIKGNANDIGMDYDALLELK--PHKRFKIGVSPAKOLPG-----GR 276  
Dh 475 --FJHEGYTHG---AGPDNDIALIKLNKYTININPICL--PRKRAASIMKTDYVGT 526  
Oy 277 IHFSGYDNDRPG---NLVYRFDVCKDETYDLYQOC-----DAQPGA----- 315  
Dh 527 V--AGWGLTQGFARNLNMFDPIDV-----HOKCATATKQYPPAKAYVNLACGL 578  
316 -----SSGYYVYVMMKROQOKERRKIIGFSGHQMVMN-GSPDQFNVAVRIT 362  
Dh 579 DRGKDCRGDSGALVFLDNETGRW--FVGIGVS---WGSINCGSGBOGYVTKVT 630

RESULT 13  
O9XV61 PRELIMINARY; PRT; 259 AA.  
AC O9XV61;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE TRIPSTN-LIKE SERINE PROTEASE (FRAGMENT).  
GN Sp-6.  
OS Ctenocephalides felis (Cat flea).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Siphonaptera; Pulicidae;  
OC Ctenocephalides.  
OC NCBI\_TaxID=7515;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-99124360; PubMed-9927170;  
RA Gaines P.J., Sampson C.M., Rushlow K.E., Stiegler G.L.;  
RT \*Cloning of a family of serine protease genes from the cat flea  
RT Ctenocephalides felis.\*  
RL Insect Mol. Biol. 81:11-22(1999).  
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
EMBL: AF053919; AAD1839.1; -  
HSSP: P00763; IDPO.  
DR InterPro: IPR001254; -  
DR InterPro: IPR001314; -  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
DR SMART: SM00020; tryp-Spc; 1.  
KW Hydrolase; Protease; Serine protease.  
FT NON\_TER 1  
SO SEQUENCE 259 AA; 28299 MW; 98222E7EB430472A CRC64;

Query Match 5.9%; Score 122; DB 5; Length 259;  
Best Local Similarity 24.1%; Pred. No. 0.0034;  
Matches 49; Conservative 30; Mismatches 68; Indels 56; Gaps 10;

Oy 101 VGIYILSS--SGDGAQHDSGSSGSKRRQIYGYDSRFSIFGKDFLLNTPSTSVKLT 158  
Dh 11 VGLSAVSYYKTKKDGIVGGQADAIK---YGTQASLYQVNEHF----- 53  
Oy 159 GCTGVLAEKHLVTAHCIDHGTQYVKTQKLVGFLPKPKDGRGANDSTSAPEQMK 218  
Dh 54 -CGASILNNYIVTAHACIYDEFTY-----SVRVG---TSFQ-GARGSVHVAQIILKIPA 103  
Oy 219 FQWIRVKTHTVPKGIKGNANDIGMDYDALLELK--PHKRFKIGVSPAKOLPG 275

Dh 104 Y-----GNVTDI--DMEXALIKVRRPRLNRTVTRVTKLDVGDMPDG 145  
Oy 276 RHFF-----SGYDNDRGNLIV 292  
Dh 146 ELATVTCWGNLGEDEDDPEQLOV 168

RESULT 14  
O9VDU8 PRELIMINARY; PRT; 449 AA.  
AC O9VDU8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, last sequence update)  
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)  
DE CG7432. PROTEIN.  
GN CG7432.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OC NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE-20196006; PubMed-10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer J.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,  
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durrin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,  
RA Palazzolo M., Pittman G.S., Pan S., Saunders R.D.C., Scheeler F., Shen H.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svyrkas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.-N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT \*The genome sequence of Drosophila melanogaster.\*;  
RL Science 287:2185-2195(2000).  
CC -1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (S1).  
EMBL: AE003727; AAE55692.2; -  
DR Flybase: FBgn0038727; CG7432.  
DR InterPro: IPR001254; -  
DR InterPro: IPR001314; -  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.  
DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.

Thu Jul 12 15:25:02 2001

us-09-554-933-3.rspt

Page 8

DR PROSITE; PS00135; TRYPSIN\_SER; 1.  
K0 Hydrolase; Serine protease.  
SQ SEQUENCE 449 AA; 49331 MM; 90DB193AA10D7331 CRC64

Query Match	5.98;	Score 122;	DB 5;	Length 445;
Best Local Similarity	23.68;	Pred. No. 0.0072;		
Matches 62; Conservative	32;	Mismatches 81;	Indels 88;	Gaps 13

```

QY      160 CTGTLVAEKHLVPAHCILHG--KTVKTKQTLKRVGLKPFKDKGGAGNDSTISAMPEOM 21
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      232 CGGSLIGKRYLITLHAHCTROSKRPRAAGFYVRLGLD-----LSTMDAPPSDP 280
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      218 KQWLIRVTRFVHPKGMVIGKANDIGMDYDALLLEKKPKKKPKMGY-----SPPAK 270
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      281 VTFPAKVEKVRTH-----EFSRIGEFYNDLAILVLDKPKRSKVIYIPVCLPKGIMRMPKE 333
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      271 QLPG-----GLIHESGYD--NDRPGNL-VYRFCDVADETDYLLEYQ----- 308
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      334 RLPGRRATVVGMTYYVYGGKSTSQORAEPLIMR-----NECCDRSYFQPINENFICAGY 388
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
      309 -----CDAQGASGSGSYVPMKROOQKREKKIIGISGHOWDMN-----SSPDQFN 356
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
      389 SDGYNVACQGDGSGPLIMR-----YDSH-WVQLGVSFSGNKKCGEPGYG 431
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY      357 VAVRITPLKQAQICVWIKGNLYD 379
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      432 VYTRVY-----EYLDWIMIRDTBD 449
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 15
063207 PRELIMINARY: PRT: 482 AA.
AC 063207:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE FACTOR X.
OS Rattus norvegicus (Rat.).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=101116;
RN [1]
RP SPOUNCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
MEDLINE=96093366; PubMed=8576539;
RA Stanton C., Ross R.P., Hutson S., Wallin R.;
RT "Evidence for competition between vitamin K-dependent gamma-
RT for intracellular processing by the vitamin K-dependent gamma-
RT carboxylase."
RT Thromb. Res. 80:63-73(1995).
-1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
-1- SIMILARITY: TO CHYMOTRYPSIN SERINE PROTEASE FAMILY (SI).
CC EMBL: X798807; CAA56202.1; -.
DR HSSP; P00742; 1XK.
DR MEROPS; S01.216.
DR InterPro: IPR000152; -
DR InterPro: IPR000294; -
DR InterPro: IPR000561; -
DR InterPro: IPR000742; -
DR InterPro: IPR001254; -
DR InterPro: IPR001314; -
DR InterPro: IPR001881; -
DR InterPro: IPR002383; -
DR InterPro; EGF_2.
DR Pfam; PF00008; EGF.
DR Pfam; PF00089; trypsin.
DR Pfam; PF00594; gla.1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR PRINTS; PR00001; GLABDOXYL.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS0186; EGF_2; 2.
DR PROSITE; PS01187; EGF_CA; 1.

```

DR PROSITE: PS00011: GIV CARBOXYLYATION: 1.  
DR PROSITE: PS00134: TRYPSIN\_HIS: UNKNOWN: 1.  
DR PROSITE: PS00135: TRYPSIN\_SER: 1.  
DR PROSITE: SM00020: TRYP\_SPE: 1.  
DR SMART: S000020: tryp. spc. 1.  
KW Calcium-binding, EGF-like domain, Glycoprotein, Hydrolase  
KW Hydrolylation: Repeat: Serine protease.  
SQ SEQUENCE 482 AA: 54265 MW: 028467883954A698 CRC64:

Query Match	5.8%;	Score 120.5;	DB 11;	Length 482;
Best Local Similarity	32.3%;	Pred. No. 0.011;		
Matches	31;	Conservative 17;	Mismatches 23;	Indels 25;
			Gaps	4;

[illegible]

Search completed: July 12, 2001, 11:44:05  
Job time: 130 sec





**THIS PAGE BLANK (USPTO)**